SCHAUM'S outlines

Problem Solved

Probability and Statistics

Third Edition

897 fully solved problems

- Concise explanations of all course concepts
- Covers Basic Probability, Sampling Theory, Regression and more

USE WITH THESE COURSES

Introduction to Probability & Statistics • Probability • Business Statistics
Basic Statistics • Beginning Statistics

Murray R. Spiegel, Ph.D. • John Schiller, Ph.D. • R. Alu Srinivasan, Ph.D.

Contents

CHAPTER 1

Basic Probability

Random Experiments

We are all familiar with the importance of experiments in science and engineering. Experimentation is useful to us because we can assume that if we perform certain experiments under very nearly identical conditions, we will arrive at results that are essentially the same. In these circumstances, we are able to control the value of the variables that affect the outcome of the experiment.

However, in some experiments, we are not able to ascertain or control the value of certain variables so that the results will vary from one performance of the experiment to the next even though most of the conditions are the same. These experiments are described as *random*. The following are some examples.

EXAMPLE 1.1 If we toss a coin, the result of the experiment is that it will either come up "tails," symbolized by T (or 0), or "heads," symbolized by H (or 1), i.e., one of the elements of the set $\{H, T\}$ (or $\{0, 1\}$).

EXAMPLE 1.2 If we toss a die, the result of the experiment is that it will come up with one of the numbers in the set $\{1, 2, 3, 4, 5, 6\}$.

EXAMPLE 1.3 If we toss a coin twice, there are four results possible, as indicated by $\{HH, HT, TH, TT\}$, i.e., both heads, heads on first and tails on second, etc.

EXAMPLE 1.4 If we are making bolts with a machine, the result of the experiment is that some may be defective. Thus when a bolt is made, it will be a member of the set {defective, nondefective}.

EXAMPLE 1.5 If an experiment consists of measuring "lifetimes" of electric light bulbs produced by a company, then the result of the experiment is a time t in hours that lies in some interval—say, $0 \leq t \leq 4000$—where we assume that no bulb lasts more than 4000 hours.

Sample Spaces

A set S that consists of all possible outcomes of a random experiment is called a *sample space*, and each outcome is called a *sample point*. Often there will be more than one sample space that can describe outcomes of an experiment, but there is usually only one that will provide the most information.

EXAMPLE 1.6 If we toss a die, one sample space, or set of all possible outcomes, is given by $\{1, 2, 3, 4, 5, 6\}$ while another is {odd, even}. It is clear, however, that the latter would not be adequate to determine, for example, whether an outcome is divisible by 3.

It is often useful to portray a sample space graphically. In such cases it is desirable to use numbers in place of letters whenever possible.

EXAMPLE 1.7 If we toss a coin twice and use 0 to represent tails and 1 to represent heads, the sample space (see Example 1.3) can be portrayed by points as in Fig. 1-1 where, for example, (0, 1) represents tails on first toss and heads on second toss, i.e., TH.

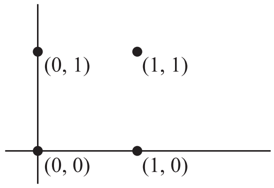


Fig. 1-1

If a sample space has a finite number of points, as in Example 1.7, it is called a *finite sample space*. If it has as many points as there are natural numbers 1, 2, 3, . . . , it is called a *countably infinite sample space*. If it has as many points as there are in some interval on the x axis, such as $0 \leq x \leq 1$, it is called a *noncountably infinite sample space*. A sample space that is finite or countably infinite is often called a *discrete sample space*, while one that is noncountably infinite is called a *nondiscrete sample space*.

Events

An *event* is a subset A of the sample space S, i.e., it is a set of possible outcomes. If the outcome of an experiment is an element of A, we say that the event A *has occurred*. An event consisting of a single point of S is often called a *simple* or *elementary event*.

EXAMPLE 1.8 If we toss a coin twice, the event that only one head comes up is the subset of the sample space that consists of points (0, 1) and (1, 0), as indicated in Fig. 1-2.

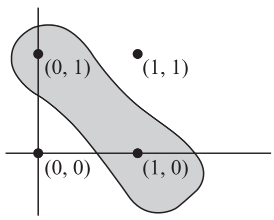


Fig. 1-2

As particular events, we have S itself, which is the *sure* or *certain event* since an element of S must occur, and the empty set $\varnothing$, which is called the *impossible event* because an element of $\varnothing$ cannot occur.

By using set operations on events in S, we can obtain other events in S. For example, if A and B are events, then

1. $A \cup B$ is the event "either A or B or both." $A \cup B$ is called the *union* of A and B.
2. $A \cap B$ is the event "both A and B." $A \cap B$ is called the *intersection* of A and B.
3. A' is the event "not A." A' is called the *complement* of A.
4. $A - B = A \cap B'$ is the event "A but not B." In particular, $A' = S - A$.

If the sets corresponding to events A and B are disjoint, i.e., $A \cap B = \varnothing$, we often say that the events are *mutually exclusive*. This means that they cannot both occur. We say that a collection of events $A_1, A_2, \ldots, A_n$ is mutually exclusive if every pair in the collection is mutually exclusive.

EXAMPLE 1.9 Referring to the experiment of tossing a coin twice, let A be the event "at least one head occurs" and B the event "the second toss results in a tail." Then $A = \{HT, TH, HH\}$, $B = \{HT, TT\}$, and so we have

$$A \cup B = \{HT, TH, HH, TT\} = S \qquad A \cap B = \{HT\}$$

$$A' = \{TT\} \qquad A - B = \{TH, HH\}$$

The Concept of Probability

In any random experiment there is always uncertainty as to whether a particular event will or will not occur. As a measure of the *chance*, or *probability*, with which we can expect the event to occur, it is convenient to assign a number between 0 and 1. If we are sure or certain that the event will occur, we say that its probability is 100% or 1, but if we are sure that the event will not occur, we say that its probability is zero. If, for example, the probability is $\frac{1}{4}$, we would say that there is a 25% chance it will occur and a 75% chance that it will not occur. Equivalently, we can say that the *odds* against its occurrence are 75% to 25%, or 3 to 1.

There are two important procedures by means of which we can estimate the probability of an event.

1. **CLASSICAL APPROACH.** If an event can occur in h different ways out of a total number of n possible ways, all of which are equally likely, then the probability of the event is h/n.

EXAMPLE 1.10 Suppose we want to know the probability that a head will turn up in a single toss of a coin. Since there are two equally likely ways in which the coin can come up—namely, heads and tails (assuming it does not roll away or stand on its edge)—and of these two ways a head can arise in only one way, we reason that the required probability is $1/2$. In arriving at this, we assume that the coin is *fair*, i.e., not *loaded* in any way.

2. **FREQUENCY APPROACH.** If after n repetitions of an experiment, where n is very large, an event is observed to occur in h of these, then the probability of the event is h/n. This is also called the *empirical probability* of the event.

EXAMPLE 1.11 If we toss a coin 1000 times and find that it comes up heads 532 times, we estimate the probability of a head coming up to be $532/1000 = 0.532$.

Both the classical and frequency approaches have serious drawbacks, the first because the words "equally likely" are vague and the second because the "large number" involved is vague. Because of these difficulties, mathematicians have been led to an *axiomatic approach* to probability.

The Axioms of Probability

Suppose we have a sample space S. If S is discrete, all subsets correspond to events and conversely, but if S is nondiscrete, only special subsets (called *measurable*) correspond to events. To each event A in the class C of events, we associate a real number $P(A)$. Then P is called a *probability function*, and $P(A)$ the *probability* of the event A, if the following axioms are satisfied.

Axiom 1 For every event A in the class C,

$$P(A) \geq 0 \tag{1}$$

Axiom 2 For the sure or certain event S in the class C,

$$P(S) = 1 \tag{2}$$

Axiom 3 For any number of mutually exclusive events $A_1, A_2, \ldots$, in the class C,

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots \tag{3}$$

In particular, for two mutually exclusive events A_1, A_2,

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) \tag{4}$$

Some Important Theorems on Probability

From the above axioms we can now prove various theorems on probability that are important in further work.

Theorem 1-1 If $A_1 \subset A_2$, then $P(A_1) \leq P(A_2)$ and $P(A_2 - A_1) = P(A_2) - P(A_1)$.

Theorem 1-2 For every event A,

$$0 \leq P(A) \leq 1, \tag{5}$$

i.e., a probability is between 0 and 1.

Theorem 1-3

$$P(\varnothing) = 0 \tag{6}$$

i.e., the impossible event has probability zero.

Theorem 1-4 If A' is the complement of A, then

$$P(A') = 1 - P(A) \tag{7}$$

Theorem 1-5 If $A = A_1 \cup A_2 \cup \cdots \cup A_n$, where $A_1, A_2, \ldots, A_n$ are mutually exclusive events, then

$$P(A) = P(A_1) + P(A_2) + \cdots + P(A_n) \tag{8}$$

In particular, if $A = S$, the sample space, then

$$P(A_1) + P(A_2) + \cdots + P(A_n) = 1 \tag{9}$$

Theorem 1-6 If A and B are any two events, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{10}$$

More generally, if A_1, A_2, A_3 are any three events, then

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) = {} & P(A_1) + P(A_2) + P(A_3) \\ & - P(A_1 \cap A_2) - P(A_2 \cap A_3) - P(A_3 \cap A_1) \\ & + P(A_1 \cap A_2 \cap A_3) \end{aligned} \tag{11}$$

Generalizations to n events can also be made.

Theorem 1-7 For any events A and B,

$$P(A) = P(A \cap B) + P(A \cap B') \tag{12}$$

Theorem 1-8 If an event A must result in the occurrence of one of the mutually exclusive events $A_1, A_2, \ldots, A_n$, then

$$P(A) = P(A \cap A_1) + P(A \cap A_2) + \cdots + P(A \cap A_n) \tag{13}$$

Assignment of Probabilities

If a sample space S consists of a finite number of outcomes $a_1, a_2, \ldots, a_n$, then by Theorem 1-5,

$$P(A_1) + P(A_2) + \cdots + P(A_n) = 1 \tag{14}$$

where $A_1, A_2, \ldots, A_n$ are elementary events given by $A_i = \{a_i\}$.

It follows that we can arbitrarily choose any nonnegative numbers for the probabilities of these simple events as long as (14) is satisfied. In particular, if we assume *equal probabilities* for all simple events, then

$$P(A_k) = \frac{1}{n}, \quad k = 1, 2, \ldots, n \tag{15}$$

and if A is any event made up of h such simple events, we have

$$P(A) = \frac{h}{n} \tag{16}$$

This is equivalent to the classical approach to probability given on page 5. We could of course use other procedures for assigning probabilities, such as the frequency approach of page 5.

Assigning probabilities provides a *mathematical model*, the success of which must be tested by experiment in much the same manner that theories in physics or other sciences must be tested by experiment.

EXAMPLE 1.12 A single die is tossed once. Find the probability of a 2 or 5 turning up.

The sample space is $S = \{1, 2, 3, 4, 5, 6\}$. If we assign equal probabilities to the sample points, i.e., if we assume that the die is fair, then

$$P(1) = P(2) = \cdots = P(6) = \frac{1}{6}$$

The event that either 2 or 5 turns up is indicated by $2 \cup 5$. Therefore,

$$P(2 \cup 5) = P(2) + P(5) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Conditional Probability

Let A and B be two events (Fig. 1-3) such that $P(A) > 0$. Denote by $P(B \mid A)$ the probability of B *given that* A has occurred. Since A is known to have occurred, it becomes the new sample space replacing the original S. From this we are led to the definition

$$P(B \mid A) \equiv \frac{P(A \cap B)}{P(A)} \tag{17}$$

or

$$P(A \cap B) \equiv P(A)\, P(B \mid A) \tag{18}$$

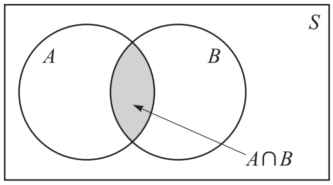


Fig. 1-3

In words, (18) says that the probability that both A and B occur is equal to the probability that A occurs times the probability that B occurs given that A has occurred. We call $P(B \mid A)$ the *conditional probability* of B given A, i.e., the probability that B will occur given that A has occurred. It is easy to show that conditional probability satisfies the axioms on page 5.

EXAMPLE 1.13 Find the probability that a single toss of a die will result in a number less than 4 if (a) no other information is given and (b) it is given that the toss resulted in an odd number.

(a) Let B denote the event {less than 4}. Since B is the union of the events 1, 2, or 3 turning up, we see by Theorem 1-5 that

$$P(B) = P(1) + P(2) + P(3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

assuming equal probabilities for the sample points.

(b) Letting A be the event {odd number}, we see that $P(A) = \frac{3}{6} = \frac{1}{2}$. Also $P(A \cap B) = \frac{2}{6} = \frac{1}{3}$. Then

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{1/3}{1/2} = \frac{2}{3}$$

Hence, the added knowledge that the toss results in an odd number raises the probability from 1/2 to 2/3.

Theorems on Conditional Probability

Theorem 1-9 For any three events A_1, A_2, A_3, we have

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_1 \cap A_2) \tag{19}$$

In words, the probability that A_1 and A_2 and A_3 all occur is equal to the probability that A_1 occurs times the probability that A_2 occurs given that A_1 has occurred times the probability that A_3 occurs given that both A_1 and A_2 have occurred. The result is easily generalized to n events.

Theorem 1-10 If an event A must result in one of the mutually exclusive events $A_1, A_2, \ldots, A_n$, then

$$P(A) = P(A_1)\, P(A \mid A_1) + P(A_2)\, P(A \mid A_2) + \cdots + P(A_n)\, P(A \mid A_n) \tag{20}$$

Independent Events

If $P(B \mid A) = P(B)$, i.e., the probability of B occurring is not affected by the occurrence or non-occurrence of A, then we say that A and B are *independent events*. This is equivalent to

$$P(A \cap B) = P(A)P(B) \tag{21}$$

as seen from (18). Conversely, if (21) holds, then A and B are independent.

We say that three events A_1, A_2, A_3 are *independent* if they are pairwise independent:

$$P(A_j \cap A_k) = P(A_j)P(A_k) \qquad j \neq k \quad \text{where} \quad j, k = 1, 2, 3 \tag{22}$$

and
$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) \tag{23}$$

Note that neither (22) nor (23) is by itself sufficient. Independence of more than three events is easily defined.

Bayes' Theorem or Rule

Suppose that $A_1, A_2, \ldots, A_n$ are mutually exclusive events whose union is the sample space S, i.e., one of the events must occur. Then if A is any event, we have the following important theorem:

Theorem 1-11 **(Bayes' Rule):**

$$P(A_k \mid A) = \frac{P(A_k)\,P(A \mid A_k)}{\sum_{j=1}^{n} P(A_j)\,P(A \mid A_j)} \tag{24}$$

This enables us to find the probabilities of the various events $A_1, A_2, \ldots, A_n$ that can *cause* A to occur. For this reason Bayes' theorem is often referred to as a *theorem on the probability of causes*.

Combinatorial Analysis

In many cases the number of sample points in a sample space is not very large, and so direct enumeration or counting of sample points needed to obtain probabilities is not difficult. However, problems arise where direct counting becomes a practical impossibility. In such cases use is made of *combinatorial analysis*, which could also be called a *sophisticated way of counting*.

Fundamental Principle of Counting: Tree Diagrams

If one thing can be accomplished in n_1 different ways and after this a second thing can be accomplished in n_2 different ways, . . . , and finally a kth thing can be accomplished in n_k different ways, then all k things can be accomplished in the specified order in $n_1 n_2 \ldots n_k$ different ways.

EXAMPLE 1.14 If a man has 2 shirts and 4 ties, then he has $2 \cdot 4 = 8$ ways of choosing a shirt and then a tie.

A diagram, called a *tree diagram* because of its appearance (Fig. 1-4), is often used in connection with the above principle.

EXAMPLE 1.15 Letting the shirts be represented by S_1, S_2 and the ties by T_1, T_2, T_3, T_4, the various ways of choosing a shirt and then a tie are indicated in the tree diagram of Fig. 1-4.

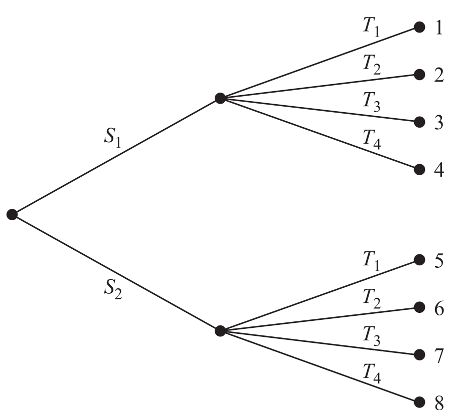


Fig. 1-4

Permutations

Suppose that we are given n distinct objects and wish to *arrange* r of these objects in a line. Since there are n ways of choosing the 1st object, and after this is done, $n - 1$ ways of choosing the 2nd object, . . . , and finally $n - r + 1$ ways of choosing the rth object, it follows by the fundamental principle of counting that the number of different *arrangements*, or *permutations* as they are often called, is given by

$$_nP_r = n(n-1)(n-2)\cdots(n-r+1) \tag{25}$$

where it is noted that the product has r factors. We call $_nP_r$ the *number of permutations of n objects taken r at a time*.

In the particular case where $r = n$, (25) becomes

$$_nP_n = n(n-1)(n-2)\cdots 1 = n! \tag{26}$$

which is called *n factorial*. We can write (25) in terms of factorials as

$$_nP_r = \frac{n!}{(n-r)!} \tag{27}$$

If $r = n$, we see that (27) and (26) agree only if we have $0! = 1$, and we shall actually take this as the definition of $0!$.

EXAMPLE 1.16 The number of different arrangements, or permutations, consisting of 3 letters each that can be formed from the 7 letters A, B, C, D, E, F, G is

$$_7P_3 = \frac{7!}{4!} = 7 \cdot 6 \cdot 5 = 210$$

Suppose that a set consists of n objects of which n_1 are of one type (i.e., indistinguishable from each other), n_2 are of a second type, . . . , n_k are of a kth type. Here, of course, $n = n_1 + n_2 + \cdots + n_k$. Then the number of different permutations of the objects is

$$_nP_{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1!n_2!\cdots n_k!} \tag{28}$$

See Problem 1.25.

EXAMPLE 1.17 The number of different permutations of the 11 letters of the word *M I S S I S S I P P I*, which consists of 1 M, 4 I's, 4 S's, and 2 P's, is

$$\frac{11!}{1!4!4!2!} = 34{,}650$$

Combinations

In a permutation we are interested in the order of arrangement of the objects. For example, *abc* is a different permutation from *bca*. In many problems, however, we are interested only in selecting or choosing objects without regard to order. Such selections are called *combinations*. For example, *abc* and *bca* are the same combination.

The total number of combinations of r objects selected from n (also called the *combinations of n things taken r at a time*) is denoted by $_nC_r$ or $\binom{n}{r}$. We have (see Problem 1.27)

$$\binom{n}{r} = {_nC_r} = \frac{n!}{r!(n-r)!} \tag{29}$$

It can also be written

$$\binom{n}{r} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{_nP_r}{r!} \tag{30}$$

It is easy to show that

$$\binom{n}{r} = \binom{n}{n-r} \qquad \text{or} \qquad {_nC_r} = {_nC_{n-r}} \tag{31}$$

EXAMPLE 1.18 The number of ways in which 3 cards can be chosen or selected from a total of 8 different cards is

$$ {}_8C_3 = \binom{8}{3} = \frac{8 \cdot 7 \cdot 6}{3!} = 56 $$

Binomial Coefficient

The numbers (29) are often called *binomial coefficients* because they arise in the *binomial expansion*

$$ (x + y)^n = x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{n}y^n \tag{32} $$

They have many interesting properties.

EXAMPLE 1.19

$$ \begin{aligned} (x + y)^4 &= x^4 + \binom{4}{1}x^3y + \binom{4}{2}x^2y^2 + \binom{4}{3}xy^3 + \binom{4}{4}y^4 \\ &= x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4 \end{aligned} $$

Stirling's Approximation to *n*!

When n is large, a direct evaluation of $n!$ may be impractical. In such cases use can be made of the approximate formula

$$ n! \sim \sqrt{2\pi n}\, n^n e^{-n} \tag{33} $$

where $e = 2.71828\ldots$, which is the base of natural logarithms. The symbol $\sim$ in (33) means that the ratio of the left side to the right side approaches 1 as $n \to \infty$.

Computing technology has largely eclipsed the value of Stirling's formula for numerical computations, but the approximation remains valuable for theoretical estimates (see Appendix A).

SOLVED PROBLEMS

Random experiments, sample spaces, and events

1.1. A card is drawn at random from an ordinary deck of 52 playing cards. Describe the sample space if consideration of suits (a) is not, (b) is, taken into account.

(a) If we do not take into account the suits, the sample space consists of ace, two, . . . , ten, jack, queen, king, and it can be indicated as $\{1, 2, \ldots, 13\}$.

(b) If we do take into account the suits, the sample space consists of ace of hearts, spades, diamonds, and clubs; . . . ; king of hearts, spades, diamonds, and clubs. Denoting hearts, spades, diamonds, and clubs, respectively, by 1, 2, 3, 4, for example, we can indicate a jack of spades by (11, 2). The sample space then consists of the 52 points shown in Fig. 1-5.

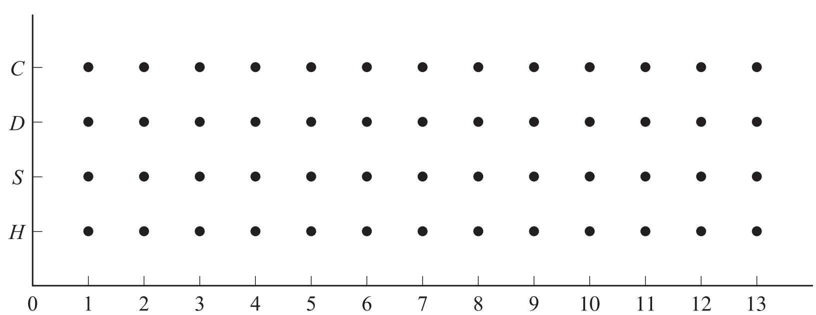


Fig. 1-5

1.2. Referring to the experiment of Problem 1.1, let A be the event {king is drawn} or simply {king} and B the event {club is drawn} or simply {club}. Describe the events (a) $A \cup B$, (b) $A \cap B$, (c) $A \cup B'$, (d) $A' \cup B'$, (e) $A - B$, (f) $A' - B'$, (g) $(A \cap B) \cup (A \cap B')$.

(a) $A \cup B$ = {either king or club (or both, i.e., king of clubs)}.

(b) $A \cap B$ = {both king and club} = {king of clubs}.

(c) Since B = {club}, B' = {not club} = {heart, diamond, spade}.

Then A $\cup$ B' = {king or heart or diamond or spade}.

(d) $A' \cup B'$ = {not king or not club} = {not king of clubs} = {any card but king of clubs}.

This can also be seen by noting that $A' \cup B' = (A \cap B)'$ and using (b).

(e) $A - B$ = {king but not club}.

This is the same as $A \cap B'$ = {king and not club}.

(f) $A' - B'$ = {not king and not "not club"} = {not king and club} = {any club except king}.

This can also be seen by noting that $A' - B' = A' \cap (B')' = A' \cap B$.

(g) $(A \cap B) \cup (A \cap B')$ = {(king and club) or (king and not club)} = {king}.

This can also be seen by noting that $(A \cap B) \cup (A \cap B') = A$.

1.3. Use Fig. 1-5 to describe the events (a) $A \cup B$, (b) $A' \cap B'$.

The required events are indicated in Fig. 1-6. In a similar manner, all the events of Problem 1.2 can also be indicated by such diagrams. It should be observed from Fig. 1-6 that $A' \cap B'$ is the complement of $A \cup B$.

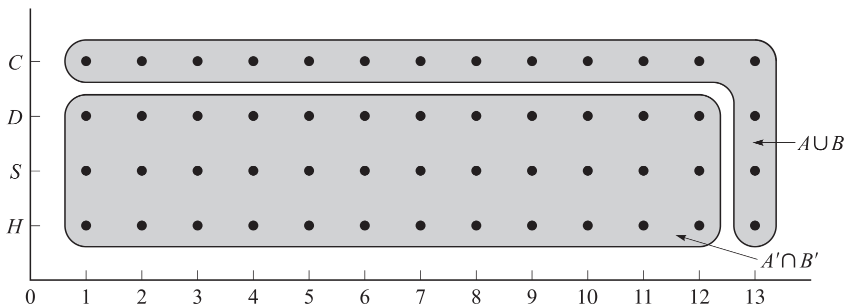


Fig. 1-6

Theorems on probability

1.4. Prove (a) Theorem 1-1, (b) Theorem 1-2, (c) Theorem 1-3, page 5.

(a) We have $A_2 = A_1 \cup (A_2 - A_1)$ where A_1 and $A_2 - A_1$ are mutually exclusive. Then by Axiom 3, page 5:

$$P(A_2) = P(A_1) + P(A_2 - A_1)$$

so that
$$P(A_2 - A_1) = P(A_2) - P(A_1)$$

Since $P(A_2 - A_1) \geq 0$ by Axiom 1, page 5, it also follows that $P(A_2) \geq P(A_1)$.

(b) We already know that $P(A) \geq 0$ by Axiom 1. To prove that $P(A) \leq 1$, we first note that A (S. Therefore, by Theorem 1-1 [part (a)] and Axiom 2,

$$P(A) \leq P(S) = 1$$

(c) We have $S = S \cup \varnothing$. Since $S \cap \varnothing = \varnothing$, it follows from Axiom 3 that

$$P(S) = P(S) + P(\varnothing) \qquad \text{or} \qquad P(\varnothing) = 0$$

1.5. Prove (a) Theorem 1-4, (b) Theorem 1-6.

(a) We have $A \cup A' = S$. Then since $A \cap A' = \varnothing$, we have

$$P(A \cup A') = P(S) \qquad \text{or} \qquad P(A) + P(A') = 1$$

i.e.,

$$P(A') = 1 - P(A)$$

(b) We have from the Venn diagram of Fig. 1-7,

(1)
$$A \cup B = A \cup [B - (A \cap B)]$$

Then since the sets A and $B - (A \cap B)$ are mutually exclusive, we have, using Axiom 3 and Theorem 1-1,

$$\begin{aligned} P(A \cup B) &= P(A) + P[B - (A \cap B)] \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

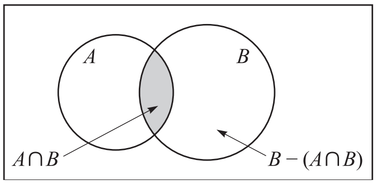


Fig. 1-7

Calculation of probabilities

1.6. A card is drawn at random from an ordinary deck of 52 playing cards. Find the probability that it is (a) an ace, (b) a jack of hearts, (c) a three of clubs or a six of diamonds, (d) a heart, (e) any suit except hearts, (f) a ten or a spade, (g) neither a four nor a club.

Let us use for brevity H, S, D, C to indicate heart, spade, diamond, club, respectively, and $1, 2, \ldots, 13$ for ace, two, . . . , king. Then $3 \cap H$ means three of hearts, while $3 \cup H$ means three or heart. Let us use the sample space of Problem 1.1(b), assigning equal probabilities of $1/52$ to each sample point. For example, $P(6 \cap C) = 1/52$.

(a)
$$\begin{aligned} P(1) &= P(1 \cap H \text{ or } 1 \cap S \text{ or } 1 \cap D \text{ or } 1 \cap C) \\ &= P(1 \cap H) + P(1 \cap S) + P(1 \cap D) + P(1 \cap C) \\ &= \frac{1}{52} + \frac{1}{52} + \frac{1}{52} + \frac{1}{52} = \frac{1}{13} \end{aligned}$$

This could also have been achieved from the sample space of Problem 1.1(a) where each sample point, in particular ace, has probability $1/13$. It could also have been arrived at by simply reasoning that there are 13 numbers and so each has probability $1/13$ of being drawn.

(b) $P(11 \cap H) = \dfrac{1}{52}$

(c) $P(3 \cap C \text{ or } 6 \cap D) = P(3 \cap C) + P(6 \cap D) = \dfrac{1}{52} + \dfrac{1}{52} = \dfrac{1}{26}$

(d) $P(H) = P(1 \cap H \text{ or } 2 \cap H \text{ or} \ldots 13 \cap H) = \dfrac{1}{52} + \dfrac{1}{52} + \cdots + \dfrac{1}{52} = \dfrac{13}{52} = \dfrac{1}{4}$

This could also have been arrived at by noting that there are four suits and each has equal probability$1/2$ of being drawn.

(e) $P(H') = 1 - P(H) = 1 - \dfrac{1}{4} = \dfrac{3}{4}$ using part (d) and Theorem 1-4, page 6.

(f) Since 10 and S are not mutually exclusive, we have, from Theorem 1-6,

$$P(10 \cup S) = P(10) + P(S) - P(10 \cap S) = \frac{1}{13} + \frac{1}{4} - \frac{1}{52} = \frac{4}{13}$$

(g) The probability of neither four nor club can be denoted by $P(4' \cap C')$. But $4' \cap C' = (4 \cup C)'$.

Therefore,

$$P(4' \cap C') = P[(4 \cup C)'] = 1 - P(4 \cup C)$$
$$= 1 - [P(4) + P(C) - P(4 \cap C)]$$
$$= 1 - \left[\frac{1}{13} + \frac{1}{4} - \frac{1}{52}\right] = \frac{9}{13}$$

We could also get this by noting that the diagram favorable to this event is the complement of the event shown circled in Fig. 1-8. Since this complement has $52 - 16 = 36$ sample points in it and each sample point is assigned probability $1/52$, the required probability is $36/52 = 9/13$.

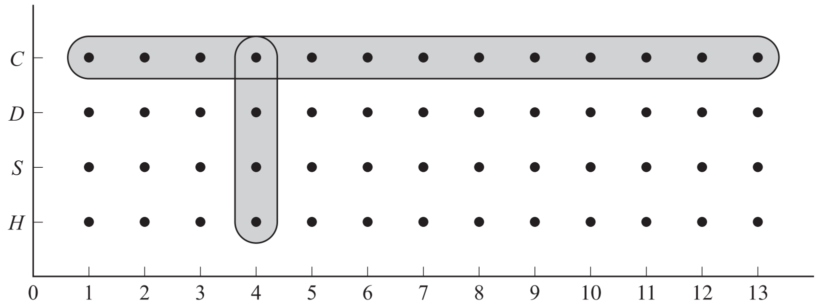


Fig. 1-8

1.7. A ball is drawn at random from a box containing 6 red balls, 4 white balls, and 5 blue balls. Determine the probability that it is (a) red, (b) white, (c) blue, (d) not red, (e) red or white.

(a) **Method 1**
Let R, W, and B denote the events of drawing a red ball, white ball, and blue ball, respectively. Then

$$P(R) = \frac{\text{ways of choosing a red ball}}{\text{total ways of choosing a ball}} = \frac{6}{6 + 4 + 5} = \frac{6}{15} = \frac{2}{5}$$

Method 2
Our sample space consists of $6 + 4 + 5 = 15$ sample points. Then if we assign equal probabilities $1/15$ to each sample point, we see that $P(R) = 6/15 = 2/5$, since there are 6 sample points corresponding to "red ball."

(b) $P(W) = \dfrac{4}{6 + 4 + 5} = \dfrac{4}{15}$

(c) $P(B) = \dfrac{5}{6 + 4 + 5} = \dfrac{5}{15} = \dfrac{1}{3}$

(d) $P(\text{not red}) = P(R') = 1 - P(R) = 1 - \dfrac{2}{5} = \dfrac{3}{5}$ by part (a).

(e) **Method 1**

$$P(\text{red or white}) = P(R \cup W) = \frac{\text{ways of choosing a red or white ball}}{\text{total ways of choosing a ball}}$$
$$= \frac{6 + 4}{6 + 4 + 5} = \frac{10}{15} = \frac{2}{3}$$

This can also be worked using the sample space as in part (a).

Method 2

$$P(R \cup W) = P(B') = 1 - P(B) = 1 - \frac{1}{3} = \frac{2}{3} \text{ by part (c).}$$

Method 3
Since events R and W are mutually exclusive, it follows from (4), page 5, that

$$P(R \cup W) = P(R) + P(W) = \frac{2}{5} + \frac{4}{15} = \frac{2}{3}$$

Conditional probability and independent events

1.8. A fair die is tossed twice. Find the probability of getting a 4, 5, or 6 on the first toss and a 1, 2, 3, or 4 on the second toss.

Let A_1 be the event "4, 5, or 6 on first toss," and A_2 be the event "1, 2, 3, or 4 on second toss." Then we are looking for $P(A_1 \cap A_2)$.

Method 1

$$P(A_1 \cap A_2) = P(A_1)P(A_2 \mid A_1) = P(A_1)P(A_2) = \left(\frac{3}{6}\right)\left(\frac{4}{6}\right) = \frac{1}{3}$$

We have used here the fact that the result of the second toss is *independent* of the first so that $P(A_2 \mid A_1) = P(A_2)$. Also we have used $P(A_1) = 3/6$ (since 4, 5, or 6 are 3 out of 6 equally likely possibilities) and $P(A_2) = 4/6$ (since 1, 2, 3, or 4 are 4 out of 6 equally likely possibilities).

Method 2

Each of the 6 ways in which a die can fall on the first toss can be associated with each of the 6 ways in which it can fall on the second toss, a total of $6 \cdot 6 = 36$ ways, all equally likely.

Each of the 3 ways in which A_1 can occur can be associated with each of the 4 ways in which A_2 can occur to give $3 \cdot 4 = 12$ ways in which both A_1 and A_2 can occur. Then

$$P(A_1 \cap A_2) = \frac{12}{36} = \frac{1}{3}$$

This shows directly that A_1 and A_2 are independent since

$$P(A_1 \cap A_2) = \frac{1}{3} = \left(\frac{3}{6}\right)\left(\frac{4}{6}\right) = P(A_1)P(A_2)$$

1.9. Find the probability of not getting a 7 or 11 total on either of two tosses of a pair of fair dice.

The sample space for each toss of the dice is shown in Fig. 1-9. For example, (5, 2) means that 5 comes up on the first die and 2 on the second. Since the dice are fair and there are 36 sample points, we assign probability $1/36$ to each.

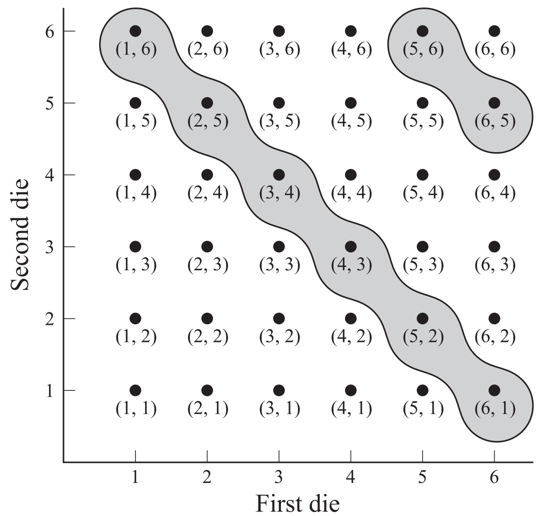


Fig. 1-9

If we let A be the event "7 or 11," then A is indicated by the circled portion in Fig. 1-9. Since 8 points are included, we have $P(A) = 8/36 = 2/9$. It follows that the probability of no 7 or 11 is given by

$$P(A') = 1 - P(A) = 1 - \frac{2}{9} = \frac{7}{9}$$

Using subscripts 1, 2 to denote 1st and 2nd tosses of the dice, we see that the probability of no 7 or 11 on either the first or second tosses is given by

$$P(A_1')P(A_2' \mid A_1') = P(A_1')P(A_2') = \left(\frac{7}{9}\right)\left(\frac{7}{9}\right) = \frac{49}{81},$$

using the fact that the tosses are independent.

1.10. Two cards are drawn from a well-shuffled ordinary deck of 52 cards. Find the probability that they are both aces if the first card is (a) replaced, (b) not replaced.

Method 1

Let A_1 = event "ace on first draw" and A_2 = event "ace on second draw." Then we are looking for $P(A_1 \cap A_2) = P(A_1)\,P(A_2 \mid A_1)$.

(a) Since for the first drawing there are 4 aces in 52 cards, $P(A_1) = 4/52$. Also, if the card is replaced for the second drawing, then $P(A_2 \mid A_1) = 4/52$, since there are also 4 aces out of 52 cards for the second drawing. Then

$$P(A_1 \cap A_2) = P(A_1)P(A_2 \mid A_1) = \left(\frac{4}{52}\right)\left(\frac{4}{52}\right) = \frac{1}{169}$$

(b) As in part (a), $P(A_1) = 4/52$. However, if an ace occurs on the first drawing, there will be only 3 aces left in the remaining 51 cards, so that $P(A_2 \mid A_1) = 3/51$. Then

$$P(A_1 \cap A_2) = P(A_1)P(A_2 \mid A_1) = \left(\frac{4}{52}\right)\left(\frac{3}{51}\right) = \frac{1}{221}$$

Method 2

(a) The first card can be drawn in any one of 52 ways, and since there is replacement, the second card can also be drawn in any one of 52 ways. Then both cards can be drawn in (52)(52) ways, all equally likely.

In such a case there are 4 ways of choosing an ace on the first draw and 4 ways of choosing an ace on the second draw so that the number of ways of choosing aces on the first and second draws is (4)(4). Then the required probability is

$$\frac{(4)(4)}{(52)(52)} = \frac{1}{169}$$

(b) The first card can be drawn in any one of 52 ways, and since there is no replacement, the second card can be drawn in any one of 51 ways. Then both cards can be drawn in (52)(51) ways, all equally likely.

In such a case there are 4 ways of choosing an ace on the first draw and 3 ways of choosing an ace on the second draw so that the number of ways of choosing aces on the first and second draws is (4)(3). Then the required probability is

$$\frac{(4)(3)}{(52)(51)} = \frac{1}{221}$$

1.11. Three balls are drawn successively from the box of Problem 1.7. Find the probability that they are drawn in the order red, white, and blue if each ball is (a) replaced, (b) not replaced.

Let R_1 = event "red on first draw," W_2 = event "white on second draw," B_3 = event "blue on third draw." We require $P(R_1 \cap W_2 \cap B_3)$.

(a) If each ball is replaced, then the events are independent and

$$\begin{aligned} P(R_1 \cap W_2 \cap B_3) &= P(R_1)P(W_2 \mid R_1)P(B_3 \mid R_2 \cap W_2) \\ &= P(R_1)P(W_2)P(B_3) \\ &= \left(\frac{6}{6+4+5}\right)\left(\frac{4}{6+4+5}\right)\left(\frac{5}{6+4+5}\right) = \frac{8}{225} \end{aligned}$$

(b) If each ball is not replaced, then the events are dependent and

$$P(R_1 \cap W_2 \cap B_3) = P(R_1)P(W_2 \mid R_1)P(B_3 \mid R_1 \cap W_2)$$

$$= \left(\frac{6}{6+4+5}\right)\left(\frac{4}{5+4+5}\right)\left(\frac{5}{5+3+5}\right) = \frac{4}{91}$$

1.12. Find the probability of a 4 turning up at least once in two tosses of a fair die.

Let A_1 = event "4 on first toss" and A_2 = event "4 on second toss." Then

$$A_1 \cup A_2 = \text{event "4 on first toss or 4 on second toss or both"}$$

$$= \text{event "at least one 4 turns up,"}$$

and we require $P(A_1 \cup A_2)$.

Method 1

Events A_1 and A_2 are not mutually exclusive, but they are independent. Hence, by (*10*) and (*21*),

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

$$= P(A_1) + P(A_2) - P(A_1)P(A_2)$$

$$= \frac{1}{6} + \frac{1}{6} - \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{11}{36}$$

Method 2

$$P(\text{at least one 4 comes up}) + P(\text{no 4 comes up}) = 1$$

Then $$P(\text{at least one 4 comes up}) = 1 - P(\text{no 4 comes up})$$

$$= 1 - P(\text{no 4 on 1st toss and no 4 on 2nd toss})$$

$$= 1 - P(A_1' \cap A_2') = 1 - P(A_1')P(A_2')$$

$$= 1 - \left(\frac{5}{6}\right)\left(\frac{5}{6}\right) = \frac{11}{36}$$

Method 3

Total number of equally likely ways in which both dice can fall $= 6 \cdot 6 = 36$.

Also Number of ways in which A_1 occurs but not $A_2 = 5$
Number of ways in which A_2 occurs but not $A_1 = 5$
Number of ways in which both A_1 and A_2 occur $= 1$

Then the number of ways in which at least one of the events A_1 or A_2 occurs $= 5 + 5 + 1 = 11$. Therefore, $P(A_1 \cup A_2) = 11/36$.

1.13. One bag contains 4 white balls and 2 black balls; another contains 3 white balls and 5 black balls. If one ball is drawn from each bag, find the probability that (a) both are white, (b) both are black, (c) one is white and one is black.

Let W_1 = event "white ball from first bag," W_2 = event "white ball from second bag."

(a) $$P(W_1 \cap W_2) = P(W_1)P(W_2 \mid W_1) = P(W_1)P(W_2) = \left(\frac{4}{4+2}\right)\left(\frac{3}{3+5}\right) = \frac{1}{4}$$

(b) $$P(W_1' \cap W_2') = P(W_1')P(W_2' \mid W_1') = P(W_1')P(W_2') = \left(\frac{2}{4+2}\right)\left(\frac{5}{3+5}\right) = \frac{5}{24}$$

(c) The required probability is

$$1 - P(W_1 \cap W_2) - P(W_1' \cap W_2') = 1 - \frac{1}{4} - \frac{5}{24} = \frac{13}{24}$$

1.14. Prove Theorem 1-10, page 7.

We prove the theorem for the case $n = 2$. Extensions to larger values of n are easily made. If event A must result in one of the two mutually exclusive events A_1, A_2, then

$$A = (A \cap A_1) \cup (A \cap A_2)$$

But $A \cap A_1$ and $A \cap A_2$ are mutually exclusive since A_1 and A_2 are. Therefore, by Axiom 3,

$$\begin{aligned} P(A) &= P(A \cap A_1) + P(A \cap A_2) \\ &= P(A_1)\,P(A \mid A_1) + P(A_2)\,P(A \mid A_2) \end{aligned}$$

using (18), page 7.

1.15. Box I contains 3 red and 2 blue marbles while Box II contains 2 red and 8 blue marbles. A fair coin is tossed. If the coin turns up heads, a marble is chosen from Box I; if it turns up tails, a marble is chosen from Box II. Find the probability that a red marble is chosen.

Let R denote the event "a red marble is chosen" while I and II denote the events that Box I and Box II are chosen, respectively. Since a red marble can result by choosing either Box I or II, we can use the results of Problem 1.14 with $A = R, A_1 = I, A_2 = II$. Therefore, the probability of choosing a red marble is

$$P(R) = P(I)P(R \mid I) + P(II)P(R \mid II) = \left(\frac{1}{2}\right)\left(\frac{3}{3+2}\right) + \left(\frac{1}{2}\right)\left(\frac{2}{2+8}\right) = \frac{2}{5}$$

Bayes' theorem

1.16. Prove Bayes' theorem (Theorem 1-11, page 8).

Since A results in one of the mutually exclusive events $A_1, A_2, \dots, A_n$, we have by Theorem 1-10 (Problem 1.14),

$$P(A) = P(A_1)P(A \mid A_1) + \cdots + P(A_n)P(A \mid A_n) = \sum_{j=1}^{n} P(A_j)P(A \mid A_j)$$

Therefore,

$$P(A_k \mid A) = \frac{P(A_k \cap A)}{P(A)} = \frac{P(A_k)P(A \mid A_k)}{\sum_{j=1}^{n} P(A_j)P(A \mid A_j)}$$

1.17. Suppose in Problem 1.15 that the one who tosses the coin does not reveal whether it has turned up heads or tails (so that the box from which a marble was chosen is not revealed) but does reveal that a red marble was chosen. What is the probability that Box I was chosen (i.e., the coin turned up heads)?

Let us use the same terminology as in Problem 1.15, i.e., $A = R, A_1 = I, A_2 = II$. We seek the probability that Box I was chosen given that a red marble is known to have been chosen. Using Bayes' rule with $n = 2$, this probability is given by

$$P(I \mid R) = \frac{P(I)P(R \mid I)}{P(I)P(R \mid I) + P(II)P(R \mid II)} = \frac{\left(\frac{1}{2}\right)\left(\frac{3}{3+2}\right)}{\left(\frac{1}{2}\right)\left(\frac{3}{3+2}\right) + \left(\frac{1}{2}\right)\left(\frac{2}{2+8}\right)} = \frac{3}{4}$$

Combinational analysis, counting, and tree diagrams

1.18. A committee of 3 members is to be formed consisting of one representative each from labor, management, and the public. If there are 3 possible representatives from labor, 2 from management, and 4 from the public, determine how many different committees can be formed using (a) the fundamental principle of counting and (b) a tree diagram.

(a) We can choose a labor representative in 3 different ways, and after this a management representative in 2 different ways. Then there are $3 \cdot 2 = 6$ different ways of choosing a labor and management representative. With each of these ways we can choose a public representative in 4 different ways. Therefore, the number of different committees that can be formed is $3 \cdot 2 \cdot 4 = 24$.

(b) Denote the 3 labor representatives by L_1, L_2, L_3; the management representatives by M_1, M_2; and the public representatives by P_1, P_2, P_3, P_4. Then the tree diagram of Fig. 1-10 shows that there are 24 different committees in all. From this tree diagram we can list all these different committees, e.g., $L_1M_1P_1$, $L_1M_1P_2$, etc.

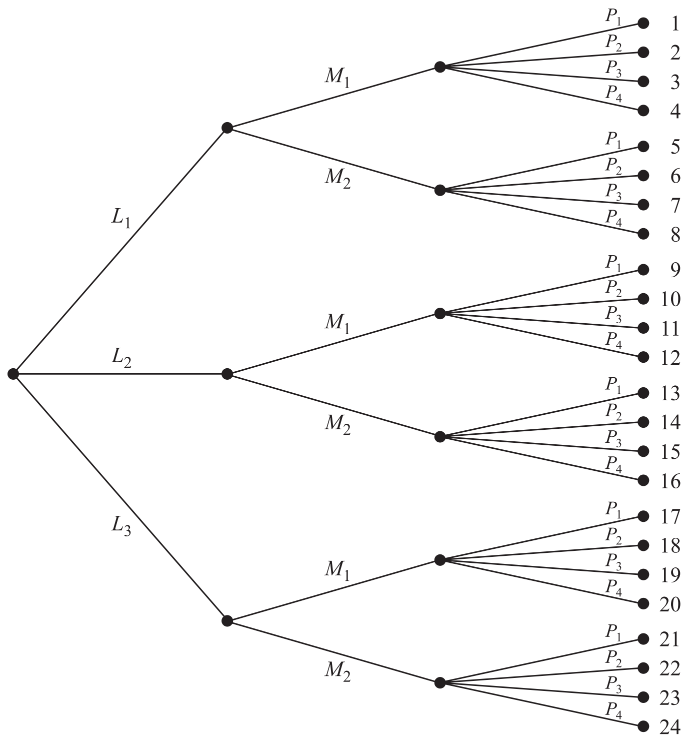


Fig. 1-10

Permutations

1.19. In how many ways can 5 differently colored marbles be arranged in a row?

We must arrange the 5 marbles in 5 positions thus: – – – – –. The first position can be occupied by any one of 5 marbles, i.e., there are 5 ways of filling the first position. When this has been done, there are 4 ways of filling the second position. Then there are 3 ways of filling the third position, 2 ways of filling the fourth position, and finally only 1 way of filling the last position. Therefore:

$$\text{Number of arrangements of 5 marbles in a row} = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5! = 120$$

In general,

$$\text{Number of arrangements of } n \text{ different objects in a row} = n(n-1)(n-2) \cdots 1 = n!$$

This is also called the *number of permutations of n different objects taken n at a time* and is denoted by ${}_nP_n$.

1.20. In how many ways can 10 people be seated on a bench if only 4 seats are available?

The first seat can be filled in any one of 10 ways, and when this has been done, there are 9 ways of filling the second seat, 8 ways of filling the third seat, and 7 ways of filling the fourth seat. Therefore:

$$\text{Number of arrangements of 10 people taken 4 at a time} = 10 \cdot 9 \cdot 8 \cdot 7 = 5040$$

In general,

$$\text{Number of arrangements of } n \text{ different objects taken } r \text{ at a time} = n(n-1) \cdots (n-r+1)$$

This is also called the *number of permutations of n different objects taken r at a time* and is denoted by ${}_nP_r$. Note that when $r = n$, ${}_nP_n = n!$ as in Problem 1.19.

1.21. Evaluate (a) ${}_8P_3$, (b) ${}_6P_4$, (c) ${}_{15}P_1$, (d) ${}_3P_3$.

(a) ${}_8P_3 = 8 \cdot 7 \cdot 6 = 336$ (b) ${}_6P_4 = 6 \cdot 5 \cdot 4 \cdot 3 = 360$ (c) ${}_{15}P_1 = 15$ (d) ${}_3P_3 = 3 \cdot 2 \cdot 1 = 6$

1.22. It is required to seat 5 men and 4 women in a row so that the women occupy the even places. How many such arrangements are possible?

The men may be seated in ${}_5P_5$ ways, and the women in ${}_4P_4$ ways. Each arrangement of the men may be associated with each arrangement of the women. Hence,

$$\text{Number of arrangements} = {}_5P_5 \cdot {}_4P_4 = 5!\,4! = (120)(24) = 2880$$

1.23. How many 4-digit numbers can be formed with the 10 digits 0, 1, 2, 3, . . . , 9 if (a) repetitions are allowed, (b) repetitions are not allowed, (c) the last digit must be zero and repetitions are not allowed?

(a) The first digit can be any one of 9 (since 0 is not allowed). The second, third, and fourth digits can be any one of 10. Then $9 \cdot 10 \cdot 10 \cdot 10 = 9000$ numbers can be formed.

(b) The first digit can be any one of 9 (any one but 0).

The second digit can be any one of 9 (any but that used for the first digit).

The third digit can be any one of 8 (any but those used for the first two digits).

The fourth digit can be any one of 7 (any but those used for the first three digits).

Then $9 \cdot 9 \cdot 8 \cdot 7 = 4536$ numbers can be formed.

Another method

The first digit can be any one of 9, and the remaining three can be chosen in ${}_9P_3$ ways. Then $9 \cdot 9P_3 = 9 \cdot 9 \cdot 8 \cdot 7 = 4536$ numbers can be formed.

(c) The first digit can be chosen in 9 ways, the second in 8 ways, and the third in 7 ways. Then $9 \cdot 8 \cdot 7 = 504$ numbers can be formed.

Another method

The first digit can be chosen in 9 ways, and the next two digits in ${}_8P_2$ ways. Then $9 \cdot {}_8P_2 = 9 \cdot 8 \cdot 7 = 504$ numbers can be formed.

1.24. Four different mathematics books, six different physics books, and two different chemistry books are to be arranged on a shelf. How many different arrangements are possible if (a) the books in each particular subject must all stand together, (b) only the mathematics books must stand together?

(a) The mathematics books can be arranged among themselves in ${}_4P_4 = 4!$ ways, the physics books in ${}_6P_6 = 6!$ ways, the chemistry books in ${}_2P_2 = 2!$ ways, and the three groups in ${}_3P_3 = 3!$ ways. Therefore,

$$\text{Number of arrangements} = 4!6!2!3! = 207{,}360.$$

(b) Consider the four mathematics books as one big book. Then we have 9 books which can be arranged in ${}_9P_9 = 9!$ ways. In all of these ways the mathematics books are together. But the mathematics books can be arranged among themselves in ${}_4P_4 = 4!$ ways. Hence,

$$\text{Number of arrangements} = 9!4! = 8{,}709{,}120$$

1.25. Five red marbles, two white marbles, and three blue marbles are arranged in a row. If all the marbles of the same color are not distinguishable from each other, how many different arrangements are possible?

Assume that there are N different arrangements. Multiplying N by the numbers of ways of arranging (a) the five red marbles among themselves, (b) the two white marbles among themselves, and (c) the three blue marbles among themselves (i.e., multiplying N by 5!2!3!), we obtain the number of ways of arranging the 10 marbles if they were all distinguishable, i.e., 10!.

Then $$(5!2!3!)N = 10! \qquad \text{and} \qquad N = 10!/(5!2!3!)$$

In general, the number of different arrangements of n objects of which n_1 are alike, n_2 are alike, . . . , n_k are alike is $\dfrac{n!}{n_1!n_2! \cdots n_k!}$ where $n_1 + n_2 + \cdots + n_k = n$.

1.26. In how many ways can 7 people be seated at a round table if (a) they can sit anywhere, (b) 2 particular people must not sit next to each other?

(a) Let 1 of them be seated anywhere. Then the remaining 6 people can be seated in $6! = 720$ ways, which is the total number of ways of arranging the 7 people in a circle.

(b) Consider the 2 particular people as 1 person. Then there are 6 people altogether and they can be arranged in 5! ways. But the 2 people considered as 1 can be arranged in 2! ways. Therefore, the number of ways of arranging 7 people at a round table with 2 particular people sitting together $= 5!2! = 240$.

Then using (a), the total number of ways in which 7 people can be seated at a round table so that the 2 particular people do not sit together $= 730 - 240 = 480$ ways.

Combinations

1.27. In how many ways can 10 objects be split into two groups containing 4 and 6 objects, respectively?

This is the same as the number of arrangements of 10 objects of which 4 objects are alike and 6 other objects are alike. By Problem 1.25, this is $\dfrac{10!}{4!6!} = \dfrac{10 \cdot 9 \cdot 8 \cdot 7}{4!} = 210$.

The problem is equivalent to finding the number of selections of 4 out of 10 objects (or 6 out of 10 objects), the order of selection being immaterial. In general, the number of selections of r out of n objects, called the *number of combinations of n things taken r at a time,* is denoted by ${}_nC_r$ or $\dbinom{n}{r}$ and is given by

$${}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{{}_nP_r}{r!}$$

1.28. Evaluate (a) ${}_7C_4$, (b) ${}_6C_5$, (c) ${}_4C_4$.

(a) ${}_7C_4 = \dfrac{7!}{4!3!} = \dfrac{7 \cdot 6 \cdot 5 \cdot 4}{4!} = \dfrac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1} = 35.$

(b) ${}_6C_5 = \dfrac{6!}{5!1!} = \dfrac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}{5!} = 6, \quad$ or $\quad {}_6C_5 = {}_6C_1 = 6.$

(c) ${}_4C_4$ is the number of selections of 4 objects taken 4 at a time, and there is only one such selection. Then ${}_4C_4 = 1$. Note that formally

$${}_4C_4 = \frac{4!}{4!0!} = 1 \quad \text{if we } \textit{define } 0! = 1.$$

1.29. In how many ways can a committee of 5 people be chosen out of 9 people?

$$\binom{9}{5} = {}_9C_5 = \frac{9!}{5!4!} = \frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5}{5!} = 126$$

1.30. Out of 5 mathematicians and 7 physicists, a committee consisting of 2 mathematicians and 3 physicists is to be formed. In how many ways can this be done if (a) any mathematician and any physicist can be included, (b) one particular physicist must be on the committee, (c) two particular mathematicians cannot be on the committee?

(a) 2 mathematicians out of 5 can be selected in ${}_5C_2$ ways.
3 physicists out of 7 can be selected in ${}_7C_3$ ways.

$$\text{Total number of possible selections} = {}_5C_2 \cdot {}_7C_3 = 10 \cdot 35 = 350$$

(b) 2 mathematicians out of 5 can be selected in ${}_5C_2$ ways.
2 physicists out of 6 can be selected in ${}_6C_2$ ways.

$$\text{Total number of possible selections} = {}_5C_2 \cdot {}_6C_2 = 10 \cdot 15 = 150$$

(c) 2 mathematicians out of 3 can be selected in ${}_3C_2$ ways.
3 physicists out of 7 can be selected in ${}_7C_3$ ways.

$$\text{Total number of possible selections} = {}_3C_2 \cdot {}_7C_3 = 3 \cdot 35 = 105$$

1.31. How many different salads can be made from lettuce, escarole, endive, watercress, and chicory?

Each green can be dealt with in 2 ways, as it can be chosen or not chosen. Since each of the 2 ways of dealing with a green is associated with 2 ways of dealing with each of the other greens, the number of ways of dealing with the 5 greens $= 2^5$ ways. But 2^5 ways includes the case in which no greens is chosen. Hence,

$$\text{Number of salads} = 2^5 - 1 = 31$$

Another method

One can select either 1 out of 5 greens, 2 out of 5 greens, . . . , 5 out of 5 greens. Then the required number of salads is

$${}_5C_1 + {}_5C_2 + {}_5C_3 + {}_5C_4 + {}_5C_5 = 5 + 10 + 10 + 5 + 1 = 31$$

In general, for any positive integer n, ${}_nC_1 + {}_nC_2 + {}_nC_3 + \cdots + {}_nC_n = 2^n - 1$.

1.32. From 7 consonants and 5 vowels, how many words can be formed consisting of 4 different consonants and 3 different vowels? The words need not have meaning.

The 4 different consonants can be selected in ${}_7C_4$ ways, the 3 different vowels can be selected in ${}_5C_3$ ways, and the resulting 7 different letters (4 consonants, 3 vowels) can then be arranged among themselves in ${}_7P_7 = 7!$ ways. Then

$$\text{Number of words} = {}_7C_4 \cdot {}_5C_3 \cdot 7! = 35 \cdot 10 \cdot 5040 = 1{,}764{,}000$$

The Binomial Coefficients

1.33. Prove that $\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$.

We have

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{n(n-1)!}{r!(n-r)!} = \frac{(n-r+r)(n-1)!}{r!(n-r)!}$$

$$= \frac{(n-r)(n-1)!}{r!(n-r)!} + \frac{r(n-1)!}{r!(n-r)!}$$

$$= \frac{(n-1)!}{r!(n-r-1)!} + \frac{(n-1)!}{(r-1)!(n-r)!}$$

$$= \binom{n-1}{r} + \binom{n-1}{r-1}$$

The result has the following interesting application. If we write out the coefficients in the binomial expansion of $(x + y)^n$ for $n = 0, 1, 2, \ldots$, we obtain the following arrangement, called *Pascal's triangle*:

$$\begin{array}{lccccccccccccc}
n = 0 & & & & & & & 1 & & & & & & \\
n = 1 & & & & & & 1 & & 1 & & & & & \\
n = 2 & & & & & 1 & & 2 & & 1 & & & & \\
n = 3 & & & & 1 & & 3 & & 3 & & 1 & & & \\
n = 4 & & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
n = 5 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
n = 6 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 \\
\text{etc.} & & & & & & & & & & & & &
\end{array}$$

An entry in any line can be obtained by adding the two entries in the preceding line that are to its immediate left and right. Therefore, $10 = 4 + 6$, $15 = 10 + 5$, etc.

1.34. Find the constant term in the expansion of $\left(x^2 + \frac{1}{x}\right)^{12}$.

According to the binomial theorem,

$$\left(x^2 + \frac{1}{x}\right)^{12} = \sum_{k=0}^{12} \binom{12}{k} (x^2)^k \left(\frac{1}{x}\right)^{12-k} = \sum_{k=0}^{12} \binom{12}{k} x^{3k-12}.$$

The constant term corresponds to the one for which $3k - 12 = 0$, i.e., $k = 4$, and is therefore given by

$$\binom{12}{4} = \frac{12 \cdot 11 \cdot 10 \cdot 9}{4 \cdot 3 \cdot 2 \cdot 1} = 495$$

Probability using combinational analysis

1.35. A box contains 8 red, 3 white, and 9 blue balls. If 3 balls are drawn at random without replacement, determine the probability that (a) all 3 are red, (b) all 3 are white, (c) 2 are red and 1 is white, (d) at least 1 is white, (e) 1 of each color is drawn, (f) the balls are drawn in the order red, white, blue.

(a) **Method 1**

Let R_1, R_2, R_3 denote the events, "red ball on 1st draw," "red ball on 2nd draw," "red ball on 3rd draw," respectively. Then $R_1 \cap R_2 \cap R_3$ denotes the event "all 3 balls drawn are red." We therefore have

$$\begin{aligned} P(R_1 \cap R_2 \cap R_3) &= P(R_1)\,P(R_2 \mid R_1)\,P(R_3 \mid R_1 \cap R_2) \\ &= \left(\frac{8}{20}\right)\left(\frac{7}{19}\right)\left(\frac{6}{18}\right) = \frac{14}{285} \end{aligned}$$

Method 2

$$\text{Required probability} = \frac{\text{number of selections of 3 out of 8 red balls}}{\text{number of selections of 3 out of 20 balls}} = \frac{{}_8C_3}{{}_{20}C_3} = \frac{14}{285}$$

(b) Using the second method indicated in part (a),

$$P(\text{all 3 are white}) = \frac{{}_3C_3}{{}_{20}C_3} = \frac{1}{1140}$$

The first method indicated in part (a) can also be used.

(c) P(2 are red and 1 is white)

$$\begin{aligned} &= \frac{\text{(selections of 2 out of 8 red balls)(selections of 1 out of 3 white balls)}}{\text{number of selections of 3 out of 20 balls}} \\ &= \frac{({}_8C_2)({}_3C_1)}{{}_{20}C_3} = \frac{7}{95} \end{aligned}$$

(d) $P(\text{none is white}) = \frac{{}_{17}C_3}{{}_{20}C_3} = \frac{34}{57}$. Then

$$P(\text{at least 1 is white}) = 1 - \frac{34}{57} = \frac{23}{57}$$

(e) $$P(\text{l of each color is drawn}) = \frac{({}_8C_1)({}_3C_1)({}_9C_1)}{{}_{20}C_3} = \frac{18}{95}$$

(f) $$\begin{aligned} P(\text{balls drawn in order red, white, blue}) &= \frac{1}{3!}P(\text{l of each color is drawn}) \\ &= \frac{1}{6}\left(\frac{18}{95}\right) = \frac{3}{95}, \text{ using (e)} \end{aligned}$$

Another method

$$\begin{aligned} P(R_1 \cap W_2 \cap B_3) &= P(R_1)\,P(W_2 \mid R_1)\,P(B_3 \mid R_1 \cap W_2) \\ &= \left(\frac{8}{20}\right)\left(\frac{3}{19}\right)\left(\frac{9}{18}\right) = \frac{3}{95} \end{aligned}$$

1.36. In the game of *poker* 5 cards are drawn from a pack of 52 well-shuffled cards. Find the probability that (a) 4 are aces, (b) 4 are aces and 1 is a king, (c) 3 are tens and 2 are jacks, (d) a nine, ten, jack, queen, king are obtained in any order, (e) 3 are of any one suit and 2 are of another, (f) at least 1 ace is obtained.

(a) $P(\text{4 aces}) = \dfrac{({}_4C_4)({}_{48}C_1)}{{}_{52}C_5} = \dfrac{1}{54{,}145}$.

(b) $P(\text{4 aces and 1 king}) = \dfrac{({}_4C_4)({}_4C_1)}{{}_{52}C_5} = \dfrac{1}{649{,}740}$.

(c) $P(\text{3 are tens and 2 are jacks}) = \dfrac{({}_4C_3)({}_4C_2)}{{}_{52}C_5} = \dfrac{1}{108{,}290}$.

(d) $P(\text{nine, ten, jack, queen, king in any order}) = \dfrac{({}_4C_1)({}_4C_1)({}_4C_1)({}_4C_1)({}_4C_1)}{{}_{52}C_5} = \dfrac{64}{162{,}435}$.

(e) $P(\text{3 of any one suit, 2 of another}) = \dfrac{(4 \cdot {}_{13}C_3)(3 \cdot {}_{13}C_2)}{{}_{52}C_5} = \dfrac{429}{4165}$,

since there are 4 ways of choosing the first suit and 3 ways of choosing the second suit.

(f) $P(\text{no ace}) = \dfrac{{}_{48}C_5}{{}_{52}C_5} = \dfrac{35{,}673}{54{,}145}$. Then $P(\text{at least one ace}) = 1 - \dfrac{35{,}673}{54{,}145} = \dfrac{18{,}472}{54{,}145}$.

1.37. Determine the probability of three 6s in 5 tosses of a fair die.

Let the tosses of the die be represented by the 5 spaces — — — — —. In each space we will have the events 6 or not 6 (6′). For example, three 6s and two not 6s can occur as 6 6 6′6 6′ or 6 6′6 6′6, etc.

Now the probability of the outcome 6 6 6′ 6 6′ is

$$P(6\ 6\ 6'\ 6\ 6') = P(6)\,P(6)\,P(6')\,P(6)\,P(6') = \frac{1}{6}\cdot\frac{1}{6}\cdot\frac{5}{6}\cdot\frac{1}{6}\cdot\frac{5}{6} = \left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$$

since we assume independence. Similarly,

$$P = \left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$$

for all other outcomes in which three 6s and two not 6s occur. But there are ${}_5C_3 = 10$ such outcomes, and these are mutually exclusive. Hence, the required probability is

$$P(6\ 6\ 6'6\ 6' \text{ or } 6\ 6'6\ 6'6 \text{ or} \ldots) = {}_5C_3\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 = \frac{5!}{3!2!}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 = \frac{125}{3888}$$

In general, if $p = P(A)$ and $q = 1 - p = P(A')$, then by using the same reasoning as given above, the probability of getting exactly x A's in n independent trials is

$${}_nC_x p^x q^{n-x} = \binom{n}{x} p^x q^{n-x}$$

1.38. A shelf has 6 mathematics books and 4 physics books. Find the probability that 3 particular mathematics books will be together.

All the books can be arranged among themselves in ${}_{10}P_{10} = 10!$ ways. Let us assume that the 3 particular mathematics books actually are replaced by 1 book. Then we have a total of 8 books that can be arranged among themselves in ${}_8P_8 = 8!$ ways. But the 3 mathematics books themselves can be arranged in ${}_3P_3 = 3!$ ways. The required probability is thus given by

$$\frac{8!\,3!}{10!} = \frac{1}{15}$$

Miscellaneous problems

1.39. A and B play 12 games of chess of which 6 are won by A, 4 are won by B, and 2 end in a draw. They agree to play a tournament consisting of 3 games. Find the probability that (a) A wins all 3 games, (b) 2 games end in a draw, (c) A and B win alternately, (d) B wins at least 1 game.

Let A_1, A_2, A_3 denote the events "A wins" in 1st, 2nd, and 3rd games, respectively, B_1, B_2, B_3 denote the events "B wins" in 1st, 2nd, and 3rd games, respectively. On the basis of their past performance (empirical probability), we shall assume that

$$P(\text{A wins any one game}) = \frac{6}{12} = \frac{1}{2}, \qquad P(\text{B wins any one game}) = \frac{4}{12} = \frac{1}{3}$$

(a) $$P(\text{A wins all 3 games}) = P(A_1 \cap A_2 \cap A_3) = P(A_1)\,P(A_2)\,P(A_3) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$$

assuming that the results of each game are independent of the results of any others. (This assumption would not be justifiable if either player were *psychologically influenced* by the other one's winning or losing.)

(b) In any one game the probability of a nondraw (i.e., either A or B wins) is $q = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$ and the probability of a draw is $p = 1 - q = \frac{1}{6}$. Then the probability of 2 draws in 3 trials is (see Problem 1.37)

$$\binom{3}{2} p^2 q^{3-2} = 3\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right) = \frac{5}{72}$$

(c) $$\begin{aligned} P(\text{A and B win alternately}) &= P(\text{A wins then B wins then A wins} \\ &\qquad \textit{or} \text{ B wins then A wins then B wins}) \\ &= P(A_1 \cap B_2 \cap A_3) + P(B_1 \cap A_2 \cap B_3) \\ &= P(A_1)P(B_2)P(A_3) + P(B_1)P(A_2)P(B_3) \\ &= \left(\frac{1}{2}\right)\left(\frac{1}{3}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{2}\right)\left(\frac{1}{3}\right) = \frac{5}{36} \end{aligned}$$

(d) $$\begin{aligned} P(\text{B wins at least one game}) &= 1 - P(\text{B wins no game}) \\ &= 1 - P(B'_1 \cap B'_2 \cap B'_3) \\ &= 1 - P(B'_1)P(B'_2)P(B'_3) \\ &= 1 - \left(\frac{2}{3}\right)\left(\frac{2}{3}\right)\left(\frac{2}{3}\right) = \frac{19}{27} \end{aligned}$$

1.40. A and B play a game in which they alternately toss a pair of dice. The one who is first to get a total of 7 wins the game. Find the probability that (a) the one who tosses first will win the game, (b) the one who tosses second will win the game.

(a) The probability of getting a 7 on a single toss of a pair of dice, assumed fair, is 1/6 as seen from Problem 1.9 and Fig. 1-9. If we suppose that A is the first to toss, then A will win in any of the following mutually exclusive cases with indicated associated probabilities:

(1) A wins on 1st toss. Probability $= \frac{1}{6}$.

(2) A loses on 1st toss, B then loses, A then wins. Probability $= \left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right)$.

(3) A loses on 1st toss, B loses, A loses, B loses, A wins. Probability $= \left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right)$.

. . . .

Then the probability that A wins is

$$\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) + \cdots$$

$$= \frac{1}{6}\left[1 + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^4 + \cdots\right] = \frac{1/6}{1 - (5/6)^2} = \frac{6}{11}$$

where we have used the result 6 of Appendix A with $x = (5/6)^2$.

(b) The probability that B wins the game is similarly

$$\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) + \cdots = \left(\frac{5}{6}\right)\left(\frac{1}{6}\right)\left[1 + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^4 + \cdots\right]$$

$$= \frac{5/36}{1 - (5/6)^2} = \frac{5}{11}$$

Therefore, we would give 6 to 5 odds that the first one to toss will win. Note that since

$$\frac{6}{11} + \frac{5}{11} = 1$$

the probability of a tie is zero. This would not be true if the game was limited. See Problem 1.100.

1.41. A machine produces a total of 12,000 bolts a day, which are on the average 3% defective. Find the probability that out of 600 bolts chosen at random, 12 will be defective.

Of the 12,000 bolts, 3%, or 360, are defective and 11,640 are not. Then:

$$\text{Required probability} = \frac{{}_{360}C_{12}\,{}_{11,640}C_{588}}{{}_{12,000}C_{600}}$$

1.42. A box contains 5 red and 4 white marbles. Two marbles are drawn successively from the box without replacement, and it is noted that the second one is white. What is the probability that the first is also white?

Method 1

If W_1, W_2 are the events "white on 1st draw," "white on 2nd draw," respectivley, we are looking for $P(W_1 \mid W_2)$. This is given by

$$P(W_1 \mid W_2) = \frac{P(W_1 \cap W_2)}{P(W_2)} = \frac{(4/9)(3/8)}{4/9} = \frac{3}{8}$$

Method 2

Since the second is known to be white, there are only 3 ways out of the remaining 8 in which the first can be white, so that the probability is $3/8$.

1.43. The probabilities that a husband and wife will be alive 20 years from now are given by 0.8 and 0.9, respectively. Find the probability that in 20 years (a) both, (b) neither, (c) at least one, will be alive.

Let H, W be the events that the husband and wife, respectively, will be alive in 20 years. Then $P(H) = 0.8$, $P(W) = 0.9$. We suppose that H and W are independent events, which may or may not be reasonable.

(a) $P(\text{both will be alive}) = P(H \cap W) = P(H)P(W) = (0.8)(0.9) = 0.72$.

(b) $P(\text{neither will be alive}) = P(H' \cap W') = P(H')\,P(W') = (0.2)(0.1) = 0.02$.

(c) $P(\text{at least one will be alive}) = 1 - P(\text{neither will be alive}) = 1 - 0.02 = 0.98$.

1.44. An inefficient secretary places n different letters into n differently addressed envelopes at random. Find the probability that at least one of the letters will arrive at the proper destination.

Let $A_1, A_2, \ldots A_n$ denote the events that the 1st, 2nd, . . . , nth letter is in the correct envelope. Then the event that at least one letter is in the correct envelope is $A_1 \cup A_2 \cup \cdots \cup A_n$, and we want to find $P(A_1 \cup A_2 \cup \cdots \cup A_n)$. From a generalization of the results (*10*) and (*11*), page 6, we have

(1) $$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum P(A_k) - \sum P(A_j \cap A_k) + \sum P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n)$$

where $\sum P(A_k)$ the sum of the probabilities of A_k from 1 to n, $\sum P(A_j \cap A_k)$ is the sum of the probabilities of $A_j \cap A_k$ with j and k from 1 to n and $k > j$, etc. We have, for example, the following:

(2) $$P(A_1) = \frac{1}{n} \quad \text{and similarly} \quad P(A_k) = \frac{1}{n}$$

since, of the n envelopes, only 1 will have the proper address. Also

(3) $$P(A_1 \cap A_2) = P(A_1)P(A_2 \mid A_1) = \left(\frac{1}{n}\right)\left(\frac{1}{n-1}\right)$$

since, if the 1st letter is in the proper envelope, then only 1 of the remaining $n - 1$ envelopes will be proper. In a similar way we find

(4) $$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) = \left(\frac{1}{n}\right)\left(\frac{1}{n-1}\right)\left(\frac{1}{n-2}\right)$$

etc., and finally

(5) $$P(A_1 \cap A_2 \cap \cdots \cap A_n) = \left(\frac{1}{n}\right)\left(\frac{1}{n-1}\right)\cdots\left(\frac{1}{1}\right) = \frac{1}{n!}$$

Now in the sum $\sum P(A_j \cap A_k)$ there are $\binom{n}{2} = {}_nC_2$ terms all having the value given by (3). Similarly in $\sum P(A_i \cap A_j \cap A_k)$, there are $\binom{n}{3} = {}_nC_3$ terms all having the value given by (4). Therefore, the required probability is

$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots \cup A_n) &= \binom{n}{1}\left(\frac{1}{n}\right) - \binom{n}{2}\left(\frac{1}{n}\right)\left(\frac{1}{n-1}\right) + \binom{n}{3}\left(\frac{1}{n}\right)\left(\frac{1}{n-1}\right)\left(\frac{1}{n-2}\right) \\ &\quad - \cdots + (-1)^{n-1}\binom{n}{n}\left(\frac{1}{n!}\right) \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-1}\frac{1}{n!} \end{aligned}$$

From calculus we know that (see Appendix A)

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

so that for $x = -1$

$$e^{-1} = 1 - \left(1 - \frac{1}{2!} + \frac{1}{3!} - \cdots\right)$$

or $$1 - \frac{1}{2!} + \frac{1}{3!} - \cdots = 1 - e^{-1}$$

It follows that if n is large, the required probability is very nearly $1 - e^{-1} = 0.6321$. This means that there is a good chance of at least 1 letter arriving at the proper destination. The result is remarkable in that the probability remains practically constant for all $n > 10$. Therefore, the probability that at least 1 letter will arrive at its proper destination is practically the same whether n is 10 or 10,000.

1.45. Find the probability that n people ($n \leq 365$) selected at random will have n different birthdays.

We assume that there are only 365 days in a year and that all birthdays are equally probable, assumptions which are not quite met in reality.

The first of the n people has of course some birthday with probability $365/365 = 1$. Then, if the second is to have a different birthday, it must occur on one of the other 364 days. Therefore, the probability that the second person has a birthday different from the first is $364/365$. Similarly the probability that the third person has a birthday different from the first two is $363/365$. Finally, the probability that the nth person has a birthday different from the others is $(365 - n + 1)/365$. We therefore have

$$\begin{aligned} P(\text{all } n \text{ birthdays are different}) &= \frac{365}{365}\cdot\frac{364}{365}\cdot\frac{363}{365}\cdots\frac{365-n+1}{365} \\ &= \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\cdots\left(1 - \frac{n-1}{365}\right) \end{aligned}$$

1.46. Determine how many people are required in Problem 1.45 to make the probability of distinct birthdays less than $1/2$.

Denoting the given probability by p and taking natural logarithms, we find

(1) $$\ln p = \ln\left(1 - \frac{1}{365}\right) + \ln\left(1 - \frac{2}{365}\right) + \cdots + \ln\left(1 - \frac{n-1}{365}\right)$$

But we know from calculus (Appendix A, formula 7) that

(2) $$\ln(1 - x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots$$

so that (1) can be written

$$\text{(3)} \qquad \ln p = -\left[\frac{1 + 2 + \cdots + (n-1)}{365}\right] - \frac{1}{2}\left[\frac{1^2 + 2^2 + \cdots + (n-1)^2}{(365)^2}\right] - \cdots$$

Using the facts that for $n = 2, 3, \ldots$ (Appendix A, formulas 1 and 2)

$$\text{(4)} \qquad 1 + 2 + \cdots + (n-1) = \frac{n(n-1)}{2}, \qquad 1^2 + 2^2 + \cdots + (n-1)^2 = \frac{n(n-1)(2n-1)}{6}$$

we obtain for (*3*)

$$\text{(5)} \qquad \ln p = -\frac{n(n-1)}{730} - \frac{n(n-1)(2n-1)}{12(365)^2} - \cdots$$

For n small compared to 365, say, $n < 30$, the second and higher terms on the right of (*5*) are negligible compared to the first term, so that a good approximation in this case is

$$\text{(6)} \qquad \text{In}\, p = \frac{n(n-1)}{730}$$

For $p = \frac{1}{2}$, $\ln p = -\ln 2 = -0.693$. Therefore, we have

$$\text{(7)} \qquad \frac{n(n-1)}{730} = 0.693 \qquad \text{or} \qquad n^2 - n - 506 = 0 \qquad \text{or} \qquad (n - 23)(n + 22) = 0$$

so that $n = 23$. Our conclusion therefore is that, if n is larger than 23, we can give better than even odds that at least 2 people will have the same birthday.

SUPPLEMENTARY PROBLEMS

Calculation of probabilities

1.47. Determine the probability p, or an estimate of it, for each of the following events:

(a) A king, ace, jack of clubs, or queen of diamonds appears in drawing a single card from a well-shuffled ordinary deck of cards.

(b) The sum 8 appears in a single toss of a pair of fair dice.

(c) A nondefective bolt will be found next if out of 600 bolts already examined, 12 were defective.

(d) A 7 or 11 comes up in a single toss of a pair of fair dice.

(e) At least 1 head appears in 3 tosses of a fair coin.

1.48. An experiment consists of drawing 3 cards in succession from a well-shuffled ordinary deck of cards. Let A_1 be the event "king on first draw," A_2 the event "king on second draw," and A_3 the event "king on third draw." State in words the meaning of each of the following:

(a) $P(A_1 \cap A'_2)$, (b) $P(A_1 \cup A_2)$, (c) $P(A'_1 \cup A'_2)$, (d) $P(A'_1 \cap A'_2 \cap A'_3)$, (e) $P[(A_1 \cap A_2) \cup (A'_2 \cap A_3)]$.

1.49. A marble is drawn at random from a box containing 10 red, 30 white, 20 blue, and 15 orange marbles. Find the probability that it is (a) orange or red, (b) not red or blue, (c) not blue, (d) white, (e) red, white, or blue.

1.50. Two marbles are drawn in succession from the box of Problem 1.49, replacement being made after each drawing. Find the probability that (a) both are white, (b) the first is red and the second is white, (c) neither is orange, (d) they are either red or white or both (red and white), (e) the second is not blue, (f) the first is orange, (g) at least one is blue, (h) at most one is red, (i) the first is white but the second is not, (j) only one is red.

1.51. Work Problem 1.50 with no replacement after each drawing.

Conditional probability and independent events

1.52. A box contains 2 red and 3 blue marbles. Find the probability that if two marbles are drawn at random (without replacement), (a) both are blue, (b) both are red, (c) one is red and one is blue.

1.53. Find the probability of drawing 3 aces at random from a deck of 52 ordinary cards if the cards are (a) replaced, (b) not replaced.

1.54. If at least one child in a family with 2 children is a boy, what is the probability that both children are boys?

1.55. Box *I* contains 3 red and 5 white balls, while Box *II* contains 4 red and 2 white balls. A ball is chosen at random from the first box and placed in the second box without observing its color. Then a ball is drawn from the second box. Find the probability that it is white.

Bayes' theorem or rule

1.56. A box contains 3 blue and 2 red marbles while another box contains 2 blue and 5 red marbles. A marble drawn at random from one of the boxes turns out to be blue. What is the probability that it came from the first box?

1.57. Each of three identical jewelry boxes has two drawers. In each drawer of the first box there is a gold watch. In each drawer of the second box there is a silver watch. In one drawer of the third box there is a gold watch while in the other there is a silver watch. If we select a box at random, open one of the drawers and find it to contain a silver watch, what is the probability that the other drawer has the gold watch?

1.58. Urn *I* has 2 white and 3 black balls; Urn *II*, 4 white and 1 black; and Urn *III*, 3 white and 4 black. An urn is selected at random and a ball drawn at random is found to be white. Find the probability that Urn *I* was selected.

Combinatorial analysis, counting, and tree diagrams

1.59. A coin is tossed 3 times. Use a tree diagram to determine the various possibilities that can arise.

1.60. Three cards are drawn at random (without replacement) from an ordinary deck of 52 cards. Find the number of ways in which one can draw (a) a diamond and a club and a heart in succession, (b) two hearts and then a club or a spade.

1.61. In how many ways can 3 different coins be placed in 2 different purses?

Permutations

1.62. Evaluate (a) ${}_4P_2$, (b) ${}_7P_5$, (c) ${}_{10}P_3$.

1.63. For what value of n is ${}_{n+1}P_3 = {}_nP_4$?

1.64. In how many ways can 5 people be seated on a sofa if there are only 3 seats available?

1.65. In how many ways can 7 books be arranged on a shelf if (a) any arrangement is possible, (b) 3 particular books must always stand together, (c) two particular books must occupy the ends?

1.66. How many numbers consisting of five different digits each can be made from the digits 1, 2, 3, . . . , 9 if (a) the numbers must be odd, (b) the first two digits of each number are even?

1.67. Solve Problem 1.66 if repetitions of the digits are allowed.

1.68. How many different three-digit numbers can be made with 3 fours, 4 twos, and 2 threes?

1.69. In how many ways can 3 men and 3 women be seated at a round table if (a) no restriction is imposed, (b) 2 particular women must not sit together, (c) each woman is to be between 2 men?

Combinations

1.70. Evaluate (a) ${}_5C_3$, (b) ${}_8C_4$, (c) ${}_{10}C_8$.

1.71. For what value of n is $3 \cdot {}_{n+1}C_3 = 7 \cdot {}_nC_2$?

1.72. In how many ways can 6 questions be selected out of 10?

1.73. How many different committees of 3 men and 4 women can be formed from 8 men and 6 women?

1.74. In how many ways can 2 men, 4 women, 3 boys, and 3 girls be selected from 6 men, 8 women, 4 boys and 5 girls if (a) no restrictions are imposed, (b) a particular man and woman must be selected?

1.75. In how many ways can a group of 10 people be divided into (a) two groups consisting of 7 and 3 people, (b) three groups consisting of 5, 3, and 2 people?

1.76. From 5 statisticians and 6 economists, a committee consisting of 3 statisticians and 2 economists is to be formed. How many different committees can be formed if (a) no restrictions are imposed, (b) 2 particular statisticians must be on the committee, (c) 1 particular economist cannot be on the committee?

1.77. Find the number of (a) combinations and (b) permutations of 4 letters each that can be made from the letters of the word *Tennessee*.

Binomial coefficients

1.78. Calculate (a) ${}_6C_3$, (b) $\binom{11}{4}$, (c) $({}_8C_2)({}_4C_3)/{}_{12}C_5$.

1.79. Expand (a) $(x + y)^6$, (b) $(x - y)^4$, (c) $(x - x^{-1})^5$, (d) $(x^2 + 2)^4$.

1.80. Find the coefficient of x in $\left(x + \frac{2}{x}\right)^9$.

Probability using combinatorial analysis

1.81. Find the probability of scoring a total of 7 points (a) once, (b) at least once, (c) twice, in 2 tosses of a pair of fair dice.

1.82. Two cards are drawn successively from an ordinary deck of 52 well-shuffled cards. Find the probability that (a) the first card is not a ten of clubs or an ace; (b) the first card is an ace but the second is not; (c) at least one card is a diamond; (d) the cards are not of the same suit; (e) not more than 1 card is a picture card (jack, queen, king); (f) the second card is not a picture card; (g) the second card is not a picture card given that the first was a picture card; (h) the cards are picture cards or spades or both.

1.83. A box contains 9 tickets numbered from 1 to 9, inclusive. If 3 tickets are drawn from the box 1 at a time, find the probability that they are alternately either odd, even, odd or even, odd, even.

1.84. The odds in favor of A winning a game of chess against B are 3:2. If 3 games are to be played, what are the odds (a) in favor of A winning at least 2 games out of the 3, (b) against A losing the first 2 games to B?

1.85. In the game of *bridge*, each of 4 players is dealt 13 cards from an ordinary well-shuffled deck of 52 cards. Find the probability that one of the players (say, the eldest) gets (a) 7 diamonds, 2 clubs, 3 hearts, and 1 spade; (b) a complete suit.

1.86. An urn contains 6 red and 8 blue marbles. Five marbles are drawn at random from it without replacement. Find the probability that 3 are red and 2 are blue.

1.87. (a) Find the probability of getting the sum 7 on at least 1 of 3 tosses of a pair of fair dice, (b) How many tosses are needed in order that the probability in (a) be greater than 0.95?

1.88. Three cards are drawn from an ordinary deck of 52 cards. Find the probability that (a) all cards are of one suit, (b) at least 2 aces are drawn.

1.89. Find the probability that a bridge player is given 13 cards of which 9 cards are of one suit.

Miscellaneous problems

1.90. A sample space consists of 3 sample points with associated probabilities given by $2p$, p^2, and $4p - 1$. Find the value of p.

1.91. How many words can be made from 5 letters if (a) all letters are different, (b) 2 letters are identical, (c) all letters are different but 2 particular letters cannot be adjacent?

1.92. Four integers are chosen at random between 0 and 9, inclusive. Find the probability that (a) they are all different, (b) not more than 2 are the same.

1.93. A pair of dice is tossed repeatedly. Find the probability that an 11 occurs for the first time on the 6th toss.

1.94. What is the least number of tosses needed in Problem 1.93 so that the probability of getting an 11 will be greater than (a) 0.5, (b) 0.95?

1.95. In a game of poker find the probability of getting (a) a *royal flush*, which consists of the ten, jack, queen, king, and ace of a single suit; (b) a *full house*, which consists of 3 cards of one face value and 2 of another (such as 3 tens and 2 jacks); (c) all different cards; (d) 4 aces.

1.96. The probability that a man will hit a target is $\frac{2}{3}$. If he shoots at the target until he hits it for the first time, find the probability that it will take him 5 shots to hit the target.

1.97. (a) A shelf contains 6 separate compartments. In how many ways can 4 indistinguishable marbles be placed in the compartments? (b) Work the problem if there are n compartments and r marbles. This type of problem arises in physics in connection with *Bose-Einstein statistics*.

1.98. (a) A shelf contains 6 separate compartments. In how many ways can 12 indistinguishable marbles be placed in the compartments so that no compartment is empty? (b) Work the problem if there are n compartments and r marbles where $r > n$. This type of problem arises in physics in connection with *Fermi-Dirac statistics*.

1.99. A poker player has cards 2, 3, 4, 6, 8. He wishes to discard the 8 and replace it by another card which he hopes will be a 5 (in which case he gets an "inside straight"). What is the probability that he will succeed assuming that the other three players together have (a) one 5, (b) two 5s, (c) three 5s, (d) no 5? Can the problem be worked if the number of 5s in the other players' hands is unknown? Explain.

1.100. Work Problem 1.40 if the game is limited to 3 tosses.

1.101. Find the probability that in a game of bridge (a) 2, (b) 3, (c) all 4 players have a complete suit.

ANSWERS TO SUPPLEMENTARY PROBLEMS

1.47. (a) 5/26 (b) 5/36 (c) 0.98 (d) 2/9 (e) 7/8

1.48. (a) Probability of king on first draw and no king on second draw.

(b) Probability of either a king on first draw or a king on second draw or both.

(c) No king on first draw or no king on second draw or both (no king on first and second draws).

(d) No king on first, second, and third draws.

(e) Probability of either king on first draw and king on second draw or no king on second draw and king on third draw.

1.49. (a) 1/3 (b) 3/5 (c) 11/15 (d) 2/5 (e) 4/5

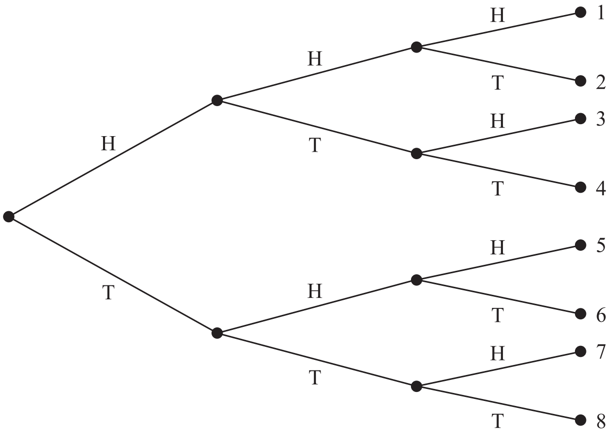
1.50. (a) 4/25 (c) 16/25 (e) 11/15 (g) 104/225 (i) 6/25
(b) 4/75 (d) 64/225 (f) 1/5 (h) 221/225 (j) 52/225

1.51. (a) 29/185 (c) 118/185 (e) 11/15 (g) 86/185 (i) 9/37
(b) 2/37 (d) 52/185 (f) 1/5 (h) 182/185 (j) 26/111

1.52. (a) 3/10 (b) 1/10 (c) 3/5 **1.53.** (a) 1/2197 (b) 1/17,576

1.54. 1/3 **1.55.** 21/56 **1.56.** 21/31 **1.57.** 1/3 **1.58.** 14/57

1.59.



1.60. (a) $13 \times 13 \times 13$ (b) $13 \times 12 \times 26$ **1.61.** 8 **1.62.** (a) 12 (b) 2520 (c) 720

1.63. $n = 5$ **1.64.** 60 **1.65.** (a) 5040 (b) 720 (c) 240 **1.66.** (a) 8400 (b) 2520

1.67. (a) 32,805 (b) 11,664 **1.68.** 26 **1.69.** (a) 120 (b) 72 (c) 12

1.70. (a) 10 (b) 70 (c) 45 **1.71.** $n = 6$ **1.72.** 210 **1.73.** 840

1.74. (a) 42,000 (b) 7000 **1.75.** (a) 120 (b) 2520 **1.76.** (a) 150 (b) 45 (c) 100

1.77. (a) 17 (b) 163 **1.78.** (a) 20 (b) 330 (c) 14/99

1.79. (a) $x^6 + 6x^5y + 15x^4y^2 + 20x^3y^3 + 15x^2y^3 + 6xy^5 + y^6$

(b) $x^4 - 4x^3y + 6x^2y^2 - 4xy^3 + y^4$

(c) $x^5 - 5x^3 + 10x - 10x^{-1} + 5x^{-3} - x^{-5}$

(d) $x^8 + 8x^6 + 24x^4 + 32x^2 + 16$

1.80. 2016 **1.81.** (a) 5/18 (b) 11/36 (c) 1/36

1.82. (a) 47/52 (b) 16/221 (c) 15/34 (d) 13/17 (e) 210/221 (f) 10/13 (g) 40/51 (h) 77/442

1.83. 5/18 **1.84.** (a) 81 : 44 (b) 21 : 4

1.85. (a) $({}_{13}C_7)({}_{13}C_2)({}_{13}C_3)({}_{13}C_1)/{}_{52}C_{13}$ (b) $4/{}_{52}C_{13}$ **1.86.** $({}_6C_3)({}_8C_2)/{}_{14}C_5$

1.87. (a) 91/216 (b) at least 17 **1.88.** (a) $4 \cdot {}_{13}C_{3//52}C_3$ (b) $({}_4C_2 \cdot {}_{48}C_1 + {}_4C_3)/{}_{52}C_3$

1.89. $4({}_{13}C_9)({}_{39}C_4)/{}_{52}C_{13}$ **1.90.** $\sqrt{11} - 3$ **1.91.** (a) 120 (b) 60 (c) 72

1.92. (a) 63/125 (b) 963/1000 **1.93.** 1,419,857/34,012,224 **1.94.** (a) 13 (b) 53

1.95. (a) $4/{}_{52}C_5$ (b) $(13)(2)(4)(6)/{}_{52}C_5$ (c) $4^5\,({}_{13}C_5)/{}_{52}C_5$ (d) (5)(4)(3)(2)/(52)(51)(50)(49)

1.96. 2/243 **1.97.** (a) 126 (b) ${}_{n+r-1}C_{n-1}$ **1.98.** (a) 462 (b) ${}_{r-1}C_{n-1}$

1.99. (a) 3/32 (b) 1/16 (c) 1/32 (d) 1/8

1.100. prob. A wins = 61/216, prob. B wins = 5/36, prob. of tie = 125/216

1.101. (a) $12/({}_{52}C_{13})({}_{39}C_{13})$ (b) $24/({}_{52}C_{13})({}_{39}C_{13})({}_{26}C_{13})$

CHAPTER 2

Random Variables and Probability Distributions

Random Variables

Suppose that to each point of a sample space we assign a number. We then have a *function* defined on the sample space. This function is called a *random variable* (or *stochastic variable*) or more precisely a *random function* (*stochastic function*). It is usually denoted by a capital letter such as X or Y. In general, a random variable has some specified physical, geometrical, or other significance.

EXAMPLE 2.1 Suppose that a coin is tossed twice so that the sample space is $S = \{HH, HT, TH, TT\}$. Let X represent the number of heads that can come up. With each sample point we can associate a number for X as shown in Table 2-1. Thus, for example, in the case of HH (i.e., 2 heads), $X = 2$ while for TH (1 head), $X = 1$. It follows that X is a random variable.

Table 2-1

Sample Point	*HH*	*HT*	*TH*	*TT*
X	2	1	1	0

It should be noted that many other random variables could also be defined on this sample space, for example, the square of the number of heads or the number of heads minus the number of tails.

A random variable that takes on a finite or countably infinite number of values (see page 4) is called a *discrete random variable* while one which takes on a noncountably infinite number of values is called a *nondiscrete random variable*.

Discrete Probability Distributions

Let X be a discrete random variable, and suppose that the possible values that it can assume are given by x_1, x_2, x_3, . . . , arranged in some order. Suppose also that these values are assumed with probabilities given by

$$P(X = x_k) = f(x_k) \qquad k = 1, 2, \ldots \tag{1}$$

It is convenient to introduce the *probability function*, also referred to as *probability distribution*, given by

$$P(X = x) = f(x) \tag{2}$$

For $x = x_k$, this reduces to (1) while for other values of x, $f(x) = 0$.

In general, $f(x)$ is a probability function if

1. $f(x) \geq 0$
2. $\sum_x f(x) = 1$

where the sum in 2 is taken over all possible values of x.

EXAMPLE 2.2 Find the probability function corresponding to the random variable X of Example 2.1. Assuming that the coin is fair, we have

$$P(HH) = \frac{1}{4} \qquad P(HT) = \frac{1}{4} \qquad P(TH) = \frac{1}{4} \qquad P(TT) = \frac{1}{4}$$

Then

$$P(X = 0) = P(TT) = \frac{1}{4}$$

$$P(X = 1) = P(HT \cup TH) = P(HT) + P(TH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$P(X = 2) = P(HH) = \frac{1}{4}$$

The probability function is thus given by Table 2-2.

Table 2-2

x	0	1	2
$f(x)$	1/4	1/2	1/4

Distribution Functions for Random Variables

The *cumulative distribution function*, or briefly the *distribution function*, for a random variable X is defined by

$$F(x) = P(X \le x) \tag{3}$$

where x is any real number, i.e., $-\infty < x < \infty$.

The distribution function $F(x)$ has the following properties:

1. $F(x)$ is nondecreasing [i.e., $F(x) \le F(y)$ if $x \le y$].
2. $\lim_{x \to -\infty} F(x) = 0$; $\lim_{x \to \infty} F(x) = 1$.
3. $F(x)$ is continuous from the right [i.e., $\lim_{h \to 0^+} F(x + h) = F(x)$ for all x].

Distribution Functions for Discrete Random Variables

The distribution function for a discrete random variable X can be obtained from its probability function by noting that, for all x in $(-\infty, \infty)$,

$$F(x) = P(X \le x) = \sum_{u \le x} f(u) \tag{4}$$

where the sum is taken over all values u taken on by X for which $u \le x$.

If X takes on only a finite number of values $x_1, x_2, \ldots, x_n$, then the distribution function is given by

$$F(x) = \begin{cases} 0 & -\infty < x < x_1 \\ f(x_1) & x_1 \le x < x_2 \\ f(x_1) + f(x_2) & x_2 \le x < x_3 \\ \vdots & \vdots \\ f(x_1) + \cdots + f(x_n) & x_n \le x < \infty \end{cases} \tag{5}$$

EXAMPLE 2.3 (a) Find the distribution function for the random variable X of Example 2.2. (b) Obtain its graph.

(a) The distribution function is

$$F(x) = \begin{cases} 0 & -\infty < x < 0 \\ \frac{1}{4} & 0 \le x < 1 \\ \frac{3}{4} & 1 \le x < 2 \\ 1 & 2 \le x < \infty \end{cases}$$

(b) The graph of $F(x)$ is shown in Fig. 2-1.

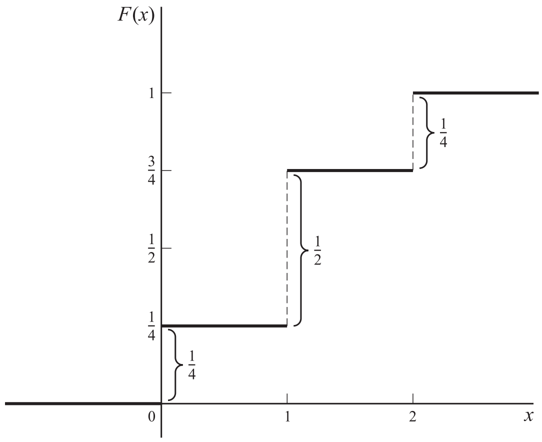


Fig. 2-1

The following things about the above distribution function, which are true in general, should be noted.

1. The magnitudes of the jumps at 0, 1, 2 are $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ which are precisely the probabilities in Table 2-2. This fact enables one to obtain the probability function from the distribution function.
2. Because of the appearance of the graph of Fig. 2-1, it is often called a *staircase function* or *step function*. The value of the function at an integer is obtained from the higher step; thus the value at 1 is $\frac{3}{4}$ and not $\frac{1}{4}$. This is expressed mathematically by stating that the distribution function is *continuous from the right* at 0, 1, 2.
3. As we proceed from left to right (i.e. going *upstairs*), the distribution function either remains the same or increases, taking on values from 0 to 1. Because of this, it is said to be a *monotonically increasing function*.

It is clear from the above remarks and the properties of distribution functions that the probability function of a discrete random variable can be obtained from the distribution function by noting that

$$f(x) = F(x) - \lim_{u \to x^-} F(u). \tag{6}$$

Continuous Random Variables

A nondiscrete random variable X is said to be *absolutely continuous*, or simply *continuous*, if its distribution function may be represented as

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du \qquad (-\infty < x < \infty) \tag{7}$$

where the function $f(x)$ has the properties

1. $f(x) \geq 0$
2. $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$

It follows from the above that if X is a continuous random variable, then the probability that X takes on any one particular value is zero, whereas the *interval probability* that X lies *between two different values*, say, a and b, is given by

$$P(a < X < b) = \int_{a}^{b} f(x)\,dx \tag{8}$$

EXAMPLE 2.4 If an individual is selected at random from a large group of adult males, the probability that his height X is precisely 68 inches (i.e., 68.000 . . . inches) would be zero. However, there is a probability greater than zero than X is between 67.000 . . . inches and 68.500 . . . inches, for example.

A function $f(x)$ that satisfies the above requirements is called a *probability function* or *probability distribution* for a continuous random variable, but it is more often called a *probability density function* or simply *density function*. Any function $f(x)$ satisfying Properties 1 and 2 above will automatically be a density function, and required probabilities can then be obtained from (8).

EXAMPLE 2.5 (a) Find the constant c such that the function

$$f(x) = \begin{cases} cx^2 & 0 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$

is a density function, and (b) compute $P(1 < X < 2)$.

(a) Since $f(x)$ satisfies Property 1 if $c \geq 0$, it must satisfy Property 2 in order to be a density function. Now

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^3 cx^2\,dx = \left.\frac{cx^3}{3}\right|_0^3 = 9c$$

and since this must equal 1, we have $c = 1/9$.

(b) $$P(1 < X < 2) = \int_1^2 \frac{1}{9}x^2\,dx = \left.\frac{x^3}{27}\right|_1^2 = \frac{8}{27} - \frac{1}{27} = \frac{7}{27}$$

In case $f(x)$ is continuous, which we shall assume unless otherwise stated, the probability that X is equal to any particular value is zero. In such case we can replace either or both of the signs $<$ in (8) by $\leq$. Thus, in Example 2.5,

$$P(1 \leq X \leq 2) = P(1 \leq X < 2) = P(1 < X \leq 2) = P(1 < X < 2) = \frac{7}{27}$$

EXAMPLE 2.6 (a) Find the distribution function for the random variable of Example 2.5. (b) Use the result of (a) to find $P(1 < x \leq 2)$.

(a) We have

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du$$

If $x < 0$, then $F(x) = 0$. If $0 \leq x < 3$, then

$$F(x) = \int_0^x f(u)\,du = \int_0^x \frac{1}{9}u^2\,du = \frac{x^3}{27}$$

If $x \geq 3$, then

$$F(x) = \int_0^3 f(u)\,du + \int_3^x f(u)\,du = \int_0^3 \frac{1}{9}u^2\,du + \int_3^x 0\,du = 1$$

Thus the required distribution function is

$$F(x) = \begin{cases} 0 & x < 0 \\ x^3/27 & 0 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

Note that $F(x)$ increases monotonically from 0 to 1 as is required for a distribution function. It should also be noted that $F(x)$ in this case is continuous.

(b) We have

$$P(1 < X \leq 2) = P(X \leq 2) - P(X \leq 1)$$
$$= F(2) - F(1)$$
$$= \frac{2^3}{27} - \frac{1^3}{27} = \frac{7}{27}$$

as in Example 2.5.

The probability that X is between x and $x + \Delta x$ is given by

$$P(x \leq X \leq x + \Delta x) = \int_x^{x+\Delta x} f(u)\,du \tag{9}$$

so that if Δx is small, we have approximately

$$P(x \leq X \leq x + \Delta x) = f(x)\Delta x \tag{10}$$

We also see from (7) on differentiating both sides that

$$\frac{dF(x)}{dx} = f(x) \tag{11}$$

at all points where $f(x)$ is continuous; i.e., the derivative of the distribution function is the density function.

It should be pointed out that random variables exist that are neither discrete nor continuous. It can be shown that the random variable X with the following distribution function is an example.

$$F(x) = \begin{cases} 0 & x < 1 \\ \dfrac{x}{2} & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

In order to obtain (11), we used the basic property

$$\frac{d}{dx}\int_a^x f(u)\,du = f(x) \tag{12}$$

which is one version of the Fundamental Theorem of Calculus.

Graphical Interpretations

If $f(x)$ is the density function for a random variable X, then we can represent $y = f(x)$ graphically by a curve as in Fig. 2-2. Since $f(x) \geq 0$, the curve cannot fall below the x axis. The entire area bounded by the curve and the x axis must be 1 because of Property 2 on page 36. Geometrically the probability that X is between a and b, i.e., $P(a < X < b)$, is then represented by the area shown shaded, in Fig. 2-2.

The distribution function $F(x) = P(X \leq x)$ is a monotonically increasing function which increases from 0 to 1 and is represented by a curve as in Fig. 2-3.

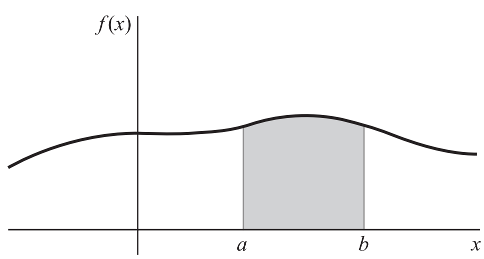


Fig. 2-2

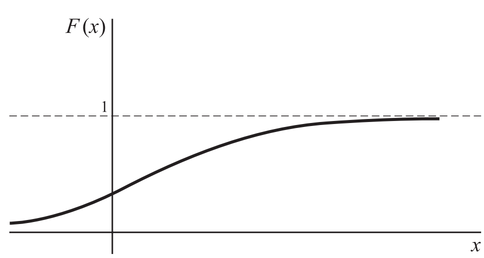


Fig. 2-3

Joint Distributions

The above ideas are easily generalized to two or more random variables. We consider the typical case of two random variables that are either both discrete or both continuous. In cases where one variable is discrete and the other continuous, appropriate modifications are easily made. Generalizations to more than two variables can also be made.

1. **DISCRETE CASE.** If X and Y are two discrete random variables, we define the *joint probability function* of X and Y by

$$P(X = x, Y = y) = f(x, y) \tag{13}$$

where 1. $f(x, y) \geq 0$

2. $\sum_x \sum_y f(x, y) = 1$

i.e., the sum over all values of x and y is 1.

Suppose that X can assume any one of m values $x_1, x_2, \ldots, x_m$ and Y can assume any one of n values $y_1, y_2, \ldots, y_n$. Then the probability of the event that $X = x_j$ and $Y = y_k$ is given by

$$P(X = x_j, Y = y_k) = f(x_j, y_k) \tag{14}$$

A joint probability function for X and Y can be represented by a *joint probability table* as in Table 2-3. The probability that $X = x_j$ is obtained by adding all entries in the row corresponding to x_i and is given by

$$P(X = x_j) = f_1(x_j) = \sum_{k=1}^{n} f(x_j, y_k) \tag{15}$$

Table 2-3

$X \backslash Y$	y_1	y_2	$\ldots$	y_n	Totals ↓	
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	$\ldots$	$f(x_1, y_n)$	$f_1(x_1)$	
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	$\ldots$	$f(x_2, y_n)$	$f_1(x_2)$	
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	
x_m	$f(x_m, y_1)$	$f(x_m, y_2)$	$\ldots$	$f(x_m, y_n)$	$f_1(x_m)$	
Totals →	$f_2(y_1)$	$f_2(y_2)$	$\ldots$	$f_2(y_n)$	1	← Grand Total

For $j = 1, 2, \ldots, m$, these are indicated by the entry totals in the extreme right-hand column or margin of Table 2-3. Similarly the probability that $Y = y_k$ is obtained by adding all entries in the column corresponding to y_k and is given by

$$P(Y = y_k) = f_2(y_k) = \sum_{j=1}^{m} f(x_j, y_k) \tag{16}$$

For $k = 1, 2, \ldots, n$, these are indicated by the entry totals in the bottom row or margin of Table 2-3.

Because the probabilities (15) and (16) are obtained from the margins of the table, we often refer to $f_1(x_j)$ and $f_2(y_k)$ [or simply $f_1(x)$ and $f_2(y)$] as the *marginal probability functions* of X and Y, respectively.

It should also be noted that

$$\sum_{j=1}^{m} f_1(x_j) = 1 \quad \sum_{k=1}^{n} f_2(y_k) = 1 \tag{17}$$

which can be written

$$\sum_{j=1}^{m} \sum_{k=1}^{n} f(x_j, y_k) = 1 \tag{18}$$

This is simply the statement that the total probability of all entries is 1. The *grand total* of 1 is indicated in the lower right-hand corner of the table.

The *joint distribution function* of X and Y is defined by

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{u \leq x} \sum_{v \leq y} f(u, v) \tag{19}$$

In Table 2-3, $F(x, y)$ is the sum of all entries for which $x_j \leq x$ and $y_k \leq y$.

2. CONTINUOUS CASE. The case where both variables are continuous is obtained easily by analogy with the discrete case on replacing sums by integrals. Thus the *joint probability function* for the random variables X and Y (or, as it is more commonly called, the *joint density function* of X and Y) is defined by

1. $f(x, y) \geq 0$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$

Graphically $z = f(x, y)$ represents a surface, called the *probability surface*, as indicated in Fig. 2-4. The total volume bounded by this surface and the xy plane is equal to 1 in accordance with Property 2 above. The probability that X lies between a and b while Y lies between c and d is given graphically by the shaded volume of Fig. 2-4 and mathematically by

$$P(a < X < b, c < Y < d) = \int_{x=a}^{b} \int_{y=c}^{d} f(x, y)\, dx\, dy \tag{20}$$

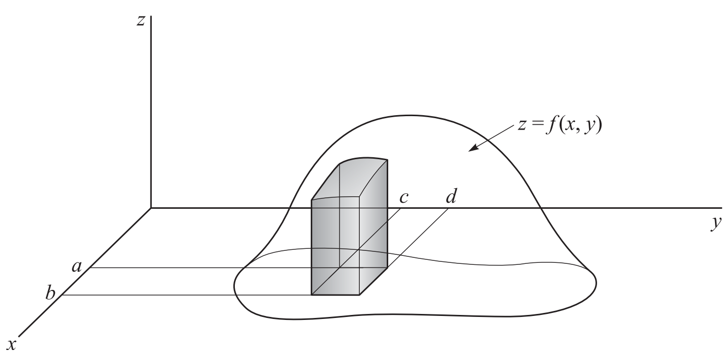


Fig. 2-4

More generally, if A represents any event, there will be a region $\mathcal{R}_A$ of the xy plane that corresponds to it. In such case we can find the probability of A by performing the integration over $\mathcal{R}_A$, i.e.,

$$P(A) = \iint_{\mathcal{R}_A} f(x, y)\, dx\, dy \tag{21}$$

The *joint distribution function* of X and Y in this case is defined by

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{u=-\infty}^{x} \int_{v=-\infty}^{y} f(u, v)\, du\, dv \tag{22}$$

It follows in analogy with (11), page 38, that

$$\frac{\partial^2 F}{\partial x\,\partial y} = f(x, y) \tag{23}$$

i.e., the density function is obtained by differentiating the distribution function with respect to x and y.

From (22) we obtain

$$P(X \leq x) = F_1(x) = \int_{u=-\infty}^{x} \int_{v=-\infty}^{\infty} f(u, v)\,du\,dv \tag{24}$$

$$P(Y \leq y) = F_2(y) = \int_{u=-\infty}^{\infty} \int_{v=-\infty}^{y} f(u, v)\,du\,dv \tag{25}$$

We call (24) and (25) the *marginal distribution functions*, or simply the *distribution functions*, of X and Y, respectively. The derivatives of (24) and (25) with respect to x and y are then called the *marginal density functions,* or simply the *density functions*, of X and Y and are given by

$$f_1(x) = \int_{v=-\infty}^{\infty} f(x, v)\,dv \qquad f_2(y) = \int_{u=-\infty}^{\infty} f(u, y)\,du \tag{26}$$

Independent Random Variables

Suppose that X and Y are discrete random variables. If the events $X = x$ and $Y = y$ are independent events for all x and y, then we say that X and Y are *independent random variables*. In such case,

$$P(X = x, Y = y) = P(X = x)P(Y = y) \tag{27}$$

or equivalently

$$f(x, y) = f_1(x)f_2(y) \tag{28}$$

Conversely, if for all x and y the joint probability function $f(x, y)$ can be expressed as the product of a function of x alone and a function of y alone (which are then the marginal probability functions of X and Y), X and Y are independent. If, however, $f(x, y)$ cannot be so expressed, then X and Y are *dependent*.

If X and Y are continuous random variables, we say that they are *independent random variables* if the events $X \leq x$ and $Y \leq y$ are independent events for all x and y. In such case we can write

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) \tag{29}$$

or equivalently

$$F(x, y) = F_1(x)F_2(y) \tag{30}$$

where $F_1(z)$ and $F_2(y)$ are the (marginal) distribution functions of X and Y, respectively. Conversely, X and Y are independent random variables if for all x and y, their joint distribution function $F(x, y)$ can be expressed as a product of a function of x alone and a function of y alone (which are the marginal distributions of X and Y, respectively). If, however, $F(x, y)$ cannot be so expressed, then X and Y are dependent.

For continuous independent random variables, it is also true that the joint density function $f(x, y)$ is the product of a function of x alone, $f_1(x)$, and a function of y alone, $f_2(y)$, and these are the (marginal) density functions of X and Y, respectively.

Change of Variables

Given the probability distributions of one or more random variables, we are often interested in finding distributions of other random variables that depend on them in some specified manner. Procedures for obtaining these distributions are presented in the following theorems for the case of discrete and continuous variables.

1. DISCRETE VARIABLES

Theorem 2-1 Let X be a discrete random variable whose probability function is $f(x)$. Suppose that a discrete random variable U is defined in terms of X by $U = \phi(X)$, where to each value of X there corresponds one and only one value of U and conversely, so that $X = \psi(U)$. Then the probability function for U is given by

$$g(u) = f[\psi(u)] \tag{31}$$

Theorem 2-2 Let X and Y be discrete random variables having joint probability function $f(x, y)$. Suppose that two discrete random variables U and V are defined in terms of X and Y by $U = \phi_1(X, Y)$, $V = \phi_2(X, Y)$, where to each pair of values of X and Y there corresponds one and only one pair of values of U and V and conversely, so that $X = \psi_1(U, V)$, $Y = \psi_2(U, V)$. Then the joint probability function of U and V is given by

$$g(u, v) = f[\psi_1(u, v), \psi_2(u, v)] \tag{32}$$

2. CONTINUOUS VARIABLES

Theorem 2-3 Let X be a continuous random variable with probability density $f(x)$. Let us define $U = \phi(X)$ where $X = \psi(U)$ as in Theorem 2-1. Then the probability density of U is given by $g(u)$ where

$$g(u)|du| = f(x)|dx| \tag{33}$$

or

$$g(u) = f(x)\left|\frac{dx}{du}\right| = f[\psi(u)]\left|\psi'(u)\right| \tag{34}$$

Theorem 2-4 Let X and Y be continuous random variables having joint density function $f(x, y)$. Let us define $U = \phi_1(X, Y)$, $V = \phi_2(X, Y)$ where $X = \psi_1(U, V)$, $Y = \psi_2(U, V)$ as in Theorem 2-2. Then the joint density function of U and V is given by $g(u, v)$ where

$$g(u, v)|du\,dv| = f(x, y)|dx\,dy| \tag{35}$$

or

$$g(u, v) = f(x, y)\left|\frac{\partial(x, y)}{\partial(u, v)}\right| = f[\psi_1(u, v), \psi_2(u, v)]\left|J\right| \tag{36}$$

In (36) the *Jacobian determinant*, or briefly *Jacobian*, is given by

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} \tag{37}$$

Probability Distributions of Functions of Random Variables

Theorems 2-2 and 2-4 specifically involve joint probability functions of two random variables. In practice one often needs to find the probability distribution of some specified function of several random variables. Either of the following theorems is often useful for this purpose.

Theorem 2-5 Let X and Y be continuous random variables and let $U = \phi_1(X, Y)$, $V = X$ (the second choice is arbitrary). Then the density function for U is the marginal density obtained from the joint density of U and V as found in Theorem 2-4. A similar result holds for probability functions of discrete variables.

Theorem 2-6 Let $f(x, y)$ be the joint density function of X and Y. Then the density function $g(u)$ of the random variable $U = \phi_1(X, Y)$ is found by differentiating with respect to u the distribution function given by

$$G(u) = P[\phi_1(X, Y) \leq u] = \iint_{\mathcal{R}} f(x, y)\,dx\,dy \tag{38}$$

Where $\mathcal{R}$ is the region for which $\phi_1(x, y) \leq u$.

Convolutions

As a particular consequence of the above theorems, we can show (see Problem 2.23) that the density function of the sum of two continuous random variables X and Y, i.e., of $U = X + Y$, having joint density function $f(x, y)$ is given by

$$g(u) = \int_{-\infty}^{\infty} f(x, u - x)\,dx \tag{39}$$

In the special case where X and Y are independent, $f(x, y) = f_1(x)f_2(y)$, and (39) reduces to

$$g(u) = \int_{-\infty}^{\infty} f_1(x)\,f_2(u - x)\,dx \tag{40}$$

which is called the *convolution* of f_1 and f_2, abbreviated, $f_1 * f_2$.

The following are some important properties of the convolution:

1. $f_1 * f_2 = f_2 * f_1$
2. $f_1 *(f_2 * f_3) = (f_1 * f_2) * f_3$
3. $f_1 *(f_2 + f_3) = f_1 * f_2 + f_1 * f_3$

These results show that f_1, f_2, f_3 obey the *commutative*, *associative*, and *distributive laws* of algebra with respect to the operation of convolution.

Conditional Distributions

We already know that if $P(A) > 0$,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \tag{41}$$

If X and Y are discrete random variables and we have the events (A: $X = x$), (B: $Y = y$), then (41) becomes

$$P(Y = y \mid X = x) = \frac{f(x, y)}{f_1(x)} \tag{42}$$

where $f(x, y) = P(X = x, Y = y)$ is the joint probability function and $f_1(x)$ is the marginal probability function for X. We define

$$f(y \mid x) \equiv \frac{f(x, y)}{f_1(x)} \tag{43}$$

and call it the *conditional probability function of Y given X*. Similarly, the conditional probability function of X given Y is

$$f(x \mid y) \equiv \frac{f(x, y)}{f_2(y)} \tag{44}$$

We shall sometimes denote $f(x \mid y)$ and $f(y \mid x)$ by $f_1(x \mid y)$ and $f_2(y \mid x)$, respectively.

These ideas are easily extended to the case where X, Y are continuous random variables. For example, the *conditional density function of Y given X* is

$$f(y \mid x) \equiv \frac{f(x, y)}{f_1(x)} \tag{45}$$

where $f(x, y)$ is the joint density function of X and Y, and $f_1(x)$ is the marginal density function of X. Using (45) we can, for example, find that the probability of Y being between c and d given that $x < X < x + dx$ is

$$P(c < Y < d \mid x < X < x + dx) = \int_c^d f(y \mid x)\, dy \tag{46}$$

Generalizations of these results are also available.

Applications to Geometric Probability

Various problems in probability arise from geometric considerations or have geometric interpretations. For example, suppose that we have a target in the form of a plane region of area K and a portion of it with area K_1, as in Fig. 2-5. Then it is reasonable to suppose that the probability of hitting the region of area K_1 is proportional to K_1. We thus define

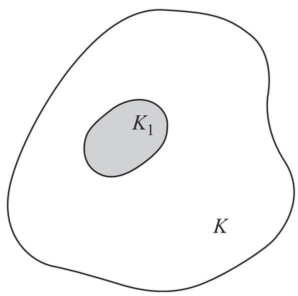


Fig. 2-5

$$P(\text{hitting region of area } K_1) = \frac{K_1}{K} \tag{47}$$

where it is assumed that the probability of hitting the target is 1. Other assumptions can of course be made. For example, there could be less probability of hitting outer areas. The type of assumption used defines the probability distribution function.

SOLVED PROBLEMS

Discrete random variables and probability distributions

2.1. Suppose that a pair of fair dice are to be tossed, and let the random variable X denote the sum of the points. Obtain the probability distribution for X.

The sample points for tosses of a pair of dice are given in Fig. 1-9, page 14. The random variable X is the sum of the coordinates for each point. Thus for (3, 2) we have $X = 5$. Using the fact that all 36 sample points are equally probable, so that each sample point has probability $1/36$, we obtain Table 2-4. For example, corresponding to $X = 5$, we have the sample points (1, 4), (2, 3), (3, 2), (4, 1), so that the associated probability is $4/36$.

Table 2-4

x	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

2.2. Find the probability distribution of boys and girls in families with 3 children, assuming equal probabilities for boys and girls.

Problem 1.37 treated the case of n mutually independent trials, where each trial had just two possible outcomes, A and A', with respective probabilities p and $q = 1 - p$. It was found that the probability of getting exactly x A's in the n trials is ${}_nC_x p^x q^{n-x}$. This result applies to the present problem, under the assumption that successive births (the "trials") are independent as far as the sex of the child is concerned. Thus, with A being the event "a boy," $n = 3$, and $p = q = \frac{1}{2}$, we have

$$P(\text{exactly } x \text{ boys}) = P(X = x) = {}_3C_x\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{3-x} = {}_3C_x\left(\frac{1}{2}\right)^3$$

where the random variable X represents the number of boys in the family. (Note that X is defined on the sample space of 3 trials.) The probability function for X,

$$f(x) = {}_3C_x\left(\frac{1}{2}\right)^3$$

is displayed in Table 2-5.

Table 2-5

x	0	1	2	3
$f(x)$	1/8	3/8	3/8	1/8

Discrete distribution functions

2.3. (a) Find the distribution function $F(x)$ for the random variable X of Problem 2.1, and (b) graph this distribution function.

(a) We have $F(x) = P(X \le x) = \sum_{u \le x} f(u)$. Then from the results of Problem 2.1, we find

$$F(x) = \begin{cases} 0 & -\infty < x < 2 \\ 1/36 & 2 \le x < 3 \\ 3/36 & 3 \le x < 4 \\ 6/36 & 4 \le x < 5 \\ \vdots & \vdots \\ 35/36 & 11 \le x < 12 \\ 1 & 12 \le x < \infty \end{cases}$$

(b) See Fig. 2-6.

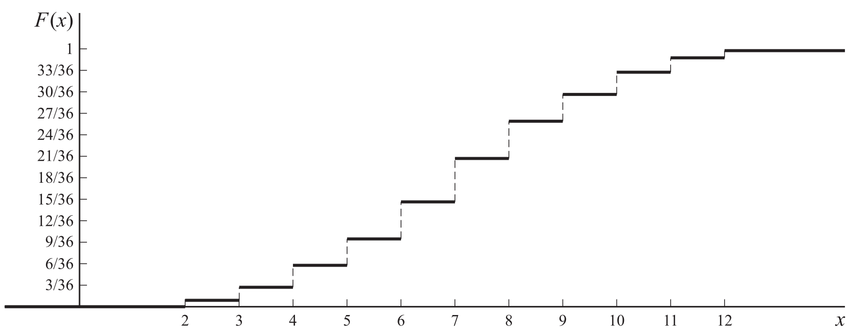


Fig. 2-6

2.4. (a) Find the distribution function $F(x)$ for the random variable X of Problem 2.2, and (b) graph this distribution function.

(a) Using Table 2-5 from Problem 2.2, we obtain

$$F(x) = \begin{cases} 0 & -\infty < x < 0 \\ 1/8 & 0 \le x < 1 \\ 1/2 & 1 \le x < 2 \\ 7/8 & 2 \le x < 3 \\ 1 & 3 \le x < \infty \end{cases}$$

(b) The graph of the distribution function of (a) is shown in Fig. 2-7.

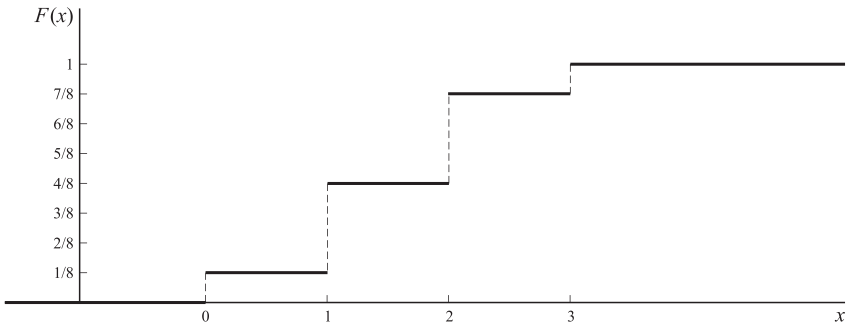


Fig. 2-7

Continuous random variables and probability distributions

2.5. A random variable X has the density function $f(x) = c/(x^2 + 1)$, where $-\infty < x < \infty$. (*a*) Find the value of the constant c. (b) Find the probability that X^2 lies between 1/3 and 1.

(a) We must have $\int_{-\infty}^{\infty} f(x)\,dx = 1$, i.e.,

$$\int_{-\infty}^{\infty} \frac{c\,dx}{x^2 + 1} = c \tan^{-1} x \Big|_{-\infty}^{\infty} = c\left[\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right] = 1$$

so that $c = 1/\pi$.

(b) If $\frac{1}{3} \le X^2 \le 1$, then either $\frac{\sqrt{3}}{3} \le X \le 1$ or $-1 \le X \le -\frac{\sqrt{3}}{3}$. Thus the required probability is

$$\frac{1}{\pi}\int_{-1}^{-\sqrt{3}/3} \frac{dx}{x^2 + 1} + \frac{1}{\pi}\int_{\sqrt{3}/3}^{1} \frac{dx}{x^2 + 1} = \frac{2}{\pi}\int_{\sqrt{3}/3}^{1} \frac{dx}{x^2 + 1}$$

$$= \frac{2}{\pi}\left[\tan^{-1}(1) - \tan^{-1}\left(\frac{\sqrt{3}}{3}\right)\right]$$

$$= \frac{2}{\pi}\left(\frac{\pi}{4} - \frac{\pi}{6}\right) = \frac{1}{6}$$

2.6. Find the distribution function corresponding to the density function of Problem 2.5.

$$F(x) = \int_{-\infty}^{x} f(u)\,du = \frac{1}{\pi}\int_{-\infty}^{x} \frac{du}{u^2 + 1} = \frac{1}{\pi}\left[\tan^{-1} u \Big|_{-\infty}^{x}\right]$$

$$= \frac{1}{\pi}[\tan^{-1} x - \tan^{-1}(-\infty)] = \frac{1}{\pi}\left[\tan^{-1} x + \frac{\pi}{2}\right]$$

$$= \frac{1}{2} + \frac{1}{\pi}\tan^{-1} x$$

2.7. The distribution function for a random variable X is

$$F(x) = \begin{cases} 1 - e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Find (a) the density function, (b) the probability that $X > 2$, and (c) the probability that $-3 < X \leq 4$.

(a) $$f(x) = \frac{d}{dx}F(x) = \begin{cases} 2e^{-2x} & x > 0 \\ 0 & x < 0 \end{cases}$$

(b) $$P(X > 2) = \int_2^{\infty} 2e^{-2u}\,du = -e^{-2u}\Big|_2^{\infty} = e^{-4}$$

Another method

By definition, $P(X \leq 2) = F(2) = 1 - e^{-4}$. Hence,

$$P(X > 2) = 1 - (1 - e^{-4}) = e^{-4}$$

(c) $$\begin{aligned} P(-3 < X \leq 4) &= \int_{-3}^{4} f(u)\,du = \int_{-3}^{0} 0\,du + \int_0^4 2e^{-2u}\,du \\ &= -e^{-2u}\Big|_0^4 = 1 - e^{-8} \end{aligned}$$

Another method

$$\begin{aligned} P(-3 < X \leq 4) &= P(X \leq 4) - P(X \leq -3) \\ &= F(4) - F(-3) \\ &= (1 - e^{-8}) - (0) = 1 - e^{-8} \end{aligned}$$

Joint distributions and independent variables

2.8. The joint probability function of two discrete random variables X and Y is given by $f(x, y) = c(2x + y)$, where x and y can assume all integers such that $0 \leq x \leq 2$, $0 \leq y \leq 3$, and $f(x, y) = 0$ otherwise.

(a) Find the value of the constant c. (c) Find $P(X \geq 1, Y \leq 2)$.
(b) Find $P(X = 2, Y = 1)$.

(a) The sample points (x, y) for which probabilities are different from zero are indicated in Fig. 2-8. The probabilities associated with these points, given by $c(2x + y)$, are shown in Table 2-6. Since the grand total, $42c$, must equal 1, we have $c = 1/42$.

Table 2-6

X \ Y	0	1	2	3	Totals ↓
0	0	c	$2c$	$3c$	$6c$
1	$2c$	$3c$	$4c$	$5c$	$14c$
2	$4c$	$5c$	$6c$	$7c$	$22c$
Totals →	$6c$	$9c$	$12c$	$15c$	$42c$

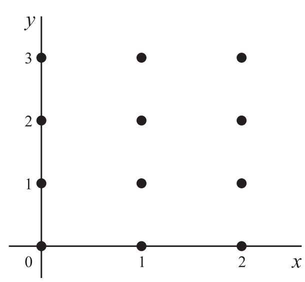


Fig. 2-8

(b) From Table 2-6 we see that

$$P(X = 2, Y = 1) = 5c + \frac{5}{42}$$

(c) From Table 2-6 we see that

$$\begin{aligned} P(X \geq 1, Y \leq 2) &= \sum_{x \geq 1} \sum_{y \leq 2} f(x, y) \\ &= (2c + 3c + 4c)(4c + 5c + 6c) \\ &= 24c = \frac{24}{42} = \frac{4}{7} \end{aligned}$$

as indicated by the entries shown shaded in the table.

2.9. Find the marginal probability functions (a) of X and (b) of Y for the random variables of Problem 2.8.

(a) The marginal probability function for X is given by $P(X = x) = f_1(x)$ and can be obtained from the margin totals in the right-hand column of Table 2-6. From these we see that

$$P(X = x) = f_1(x) = \begin{cases} 6c = 1/7 & x = 0 \\ 14c = 1/3 & x = 1 \\ 22c = 11/21 & x = 2 \end{cases}$$

Check: $\frac{1}{7} + \frac{1}{3} + \frac{11}{21} = 1$

(b) The marginal probability function for Y is given by $P(Y = y) = f_2(y)$ and can be obtained from the margin totals in the last row of Table 2-6. From these we see that

$$P(Y = y) = f_2(y) = \begin{cases} 6c = 1/7 & y = 0 \\ 9c = 3/14 & y = 1 \\ 12c = 2/7 & y = 2 \\ 15c = 5/14 & y = 3 \end{cases}$$

Check: $\frac{1}{7} + \frac{3}{14} + \frac{2}{7} + \frac{5}{14} = 1$

2.10. Show that the random variables X and Y of Problem 2.8 are dependent.

If the random variables X and Y are independent, then we must have, for all x and y,

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

But, as seen from Problems 2.8(b) and 2.9,

$$P(X = 2, Y = 1) = \frac{5}{42} \qquad P(X = 2) = \frac{11}{21} \qquad P(Y = 1) = \frac{3}{14}$$

so that

$$P(X = 2, Y = 1) \neq P(X = 2)P(Y = 1)$$

The result also follows from the fact that the joint probability function $(2x + y)/42$ cannot be expressed as a function of x alone times a function of y alone.

2.11. The joint density function of two continuous random variables X and Y is

$$f(x, y) = \begin{cases} cxy & 0 < x < 4, 1 < y < 5 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the value of the constant c.

(b) Find $P(1 < X < 2, 2 < Y < 3)$.

(c) Find $P(X \geq 3, Y \leq 2)$.

(a) We must have the total probability equal to 1, i.e.,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$$

Using the definition of $f(x, y)$, the integral has the value

$$\int_{x=0}^{4}\int_{y=1}^{5} cxy\,dx\,dy = c\int_{x=0}^{4}\left[\int_{y=1}^{5} xy\,dy\right]dx$$

$$= c\int_{z=0}^{4} \frac{xy^2}{2}\bigg|_{y=1}^{5} dx = c\int_{x=0}^{4}\left(\frac{25x}{2} - \frac{x}{2}\right)dx$$

$$= c\int_{x=0}^{4} 12x\,dx = c(6x^2)\bigg|_{x=0}^{4} = 96c$$

Then $96c = 1$ and $c = 1/96$.

(b) Using the value of c found in (a), we have

$$P(1 < X < 2, 2 < Y < 3) = \int_{x=1}^{2}\int_{y=2}^{3}\frac{xy}{96}\,dx\,dy$$

$$= \frac{1}{96}\int_{x=1}^{2}\left[\int_{y=2}^{3} xy\,dy\right]dx = \frac{1}{96}\int_{x=1}^{2}\frac{xy^2}{2}\bigg|_{y=2}^{3} dx$$

$$= \frac{1}{96}\int_{x=1}^{2}\frac{5x}{2}\,dx = \frac{5}{192}\left(\frac{x^2}{2}\right)\bigg|_{1}^{2} = \frac{5}{128}$$

(c)

$$P(X \geq 3, Y \leq 2) = \int_{x=3}^{4}\int_{y=1}^{2}\frac{xy}{96}\,dx\,dy$$

$$= \frac{1}{96}\int_{x=3}^{4}\left[\int_{y=1}^{2} xy\,dy\right]dx = \frac{1}{96}\int_{x=3}^{4}\frac{xy^2}{2}\bigg|_{y=1}^{2} dx$$

$$= \frac{1}{96}\int_{x=3}^{4}\frac{3x}{2}\,dx = \frac{7}{128}$$

2.12. Find the marginal distribution functions (a) of X and (b) of Y for Problem 2.11.

(a) The marginal distribution function for X if $0 \leq x < 4$ is

$$F_1(x) = P(X \leq x) = \int_{u=-\infty}^{x}\int_{v=-\infty}^{\infty} f(u, v)\,du\,dv$$

$$= \int_{u=0}^{x}\int_{v=1}^{5}\frac{uv}{96}\,du\,dv$$

$$= \frac{1}{96}\int_{u=0}^{x}\left[\int_{v=1}^{5} uv\,dv\right]du = \frac{x^2}{16}$$

For $x \geq 4$, $F_1(x) = 1$; for $x < 0$, $F_1(x) = 0$. Thus

$$F_1(x) = \begin{cases} 0 & x < 0 \\ x^{2/16} & 0 \leq x < 4 \\ 1 & x \geq 4 \end{cases}$$

As $F_1(x)$ is continuous at $x = 0$ and $x = 4$, we could replace $<$ by $\leq$ in the above expression.

(b) The marginal distribution function for Y if $1 \le y < 5$ is

$$F_2(y) = P(Y \le y) = \int_{u=-\infty}^{\infty} \int_{v=1}^{y} f(u, v)\, du\, dv$$

$$= \int_{u=0}^{4} \int_{v=1}^{y} \frac{uv}{96}\, du\, dv = \frac{y^2 - 1}{24}$$

For $y \ge 5$, $F_2(y) = 1$. For $y < 1$, $F_2(y) = 0$. Thus

$$F_2(y) = \begin{cases} 0 & y < 1 \\ (y^2 - 1)/24 & 1 \le y < 5 \\ 1 & y \ge 5 \end{cases}$$

As $F_2(y)$ is continuous at $y = 1$ and $y = 5$, we could replace $<$ by $\le$ in the above expression.

2.13. Find the joint distribution function for the random variables X, Y of Problem 2.11.

From Problem 2.11 it is seen that the joint density function for X and Y can be written as the product of a function of x alone and a function of y alone. In fact, $f(x, y) = f_1(x)f_2(y)$, where

$$f_1(x) = \begin{cases} c_1 x & 0 < x < 4 \\ 0 & \text{otherwise} \end{cases} \qquad f_2(y) = \begin{cases} c_2 y & 1 < y < 5 \\ 0 & \text{otherwise} \end{cases}$$

and $c_1 c_2 = c = 1/96$. It follows that X and Y are independent, so that their joint distribution function is given by $F(x, y) = F_1(x)F_2(y)$. The marginal distributions $F_1(x)$ and $F_2(y)$ were determined in Problem 2.12, and Fig. 2-9 shows the resulting piecewise definition of $F(x, y)$.

2.14. In Problem 2.11 find $P(X + Y < 3)$.

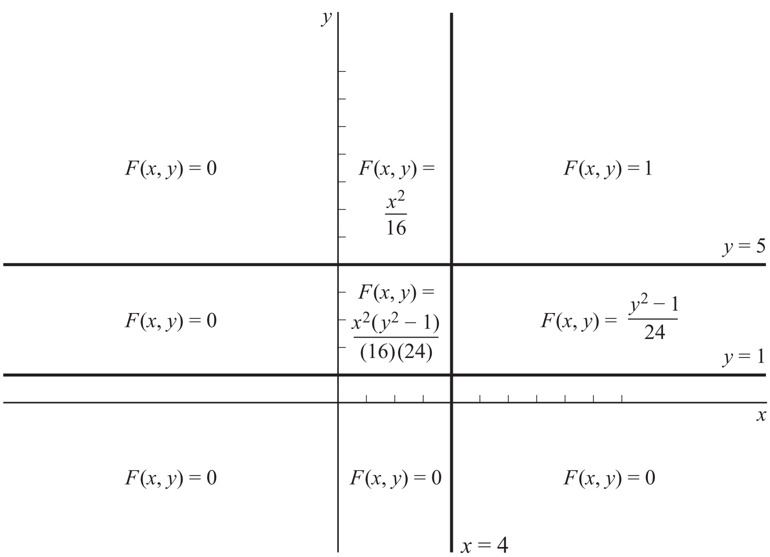


Fig. 2-9

In Fig. 2-10 we have indicated the square region $0 < x < 4$, $1 < y < 5$ within which the joint density function of X and Y is different from zero. The required probability is given by

$$P(X + Y < 3) = \iint_{\mathscr{R}} f(x, y)\, dx\, dy$$

where $\mathscr{R}$ is the part of the square over which $x + y < 3$, shown shaded in Fig. 2-10. Since $f(x, y) = xy/96$ over $\mathscr{R}$, this probability is given by

$$\int_{x=0}^{2}\int_{y=1}^{3-x}\frac{xy}{96}\,dx\,dy$$

$$= \frac{1}{96}\int_{x=0}^{2}\left[\int_{y=1}^{3-x} xy\,dy\right]dx$$

$$= \frac{1}{96}\int_{x=0}^{2}\frac{xy^2}{2}\bigg|_{y=1}^{3-x}dx = \frac{1}{192}\int_{x=0}^{2}[x(3-x)^2 - x] = \frac{1}{48}$$

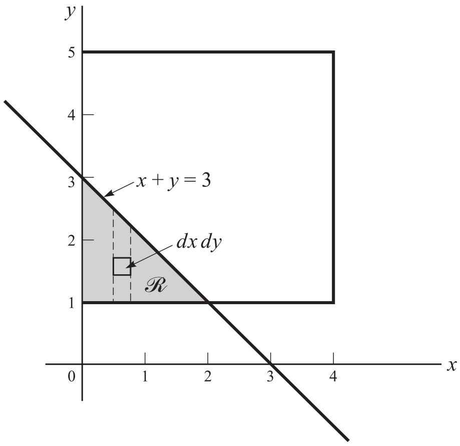


Fig. 2-10

Change of variables

2.15. Prove Theorem 2-1, page 42.

The probability function for U is given by

$$g(u) = P(U = u) = P[\phi(X) = u] = P[X = \psi(u)] = f[\psi(u)]$$

In a similar manner Theorem 2-2, page 42, can be proved.

2.16. Prove Theorem 2-3, page 42.

Consider first the case where $u = \phi(x)$ or $x = \psi(u)$ is an increasing function, i.e., u increases as x increases (Fig. 2-11). There, as is clear from the figure, we have

(1) $$P(u_1 < U < u_2) = P(x_1 < X < x_2)$$

or

(2) $$\int_{u_1}^{u_2} g(u)\,du = \int_{x_1}^{x_2} f(x)\,dx$$

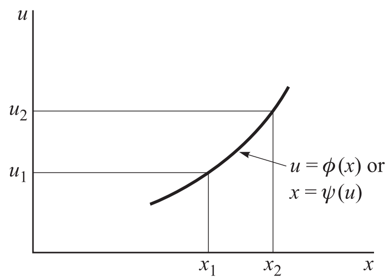


Fig. 2-11

Letting $x = \psi(u)$ in the integral on the right, (2) can be written

$$\int_{u_1}^{u_2} g(u)\,du = \int_{u_1}^{u_2} f[\psi(u)]\psi'(u)\,du$$

This can hold for all u_1 and u_2 only if the integrands are identical, i.e.,

$$g(u) = f[\psi(u)]\psi'(u)$$

This is a special case of (34), page 42, where $\psi'(u) > 0$ (i.e., the slope is positive). For the case where $\psi'(u) \le 0$, i.e., u is a decreasing function of x, we can also show that (34) holds (see Problem 2.67). The theorem can also be proved if $\psi'(u) \ge 0$ or $\psi'(u) < 0$.

2.17. Prove Theorem 2-4, page 42.

We suppose first that as x and y increase, u and v also increase. As in Problem 2.16 we can then show that

$$P(u_1 < U < u_2, v_1 < V < v_2) = P(x_1 < X < x_2, y_1 < Y < y_2)$$

or

$$\int_{v_1}^{u_2}\int_{v_1}^{v_2} g(u, v)\,du\,dv = \int_{x_1}^{x_2}\int_{y_1}^{y_2} f(x, y)\,dx\,dy$$

Letting $x = \psi_1(u, v)$, $y = \psi_2(u, v)$ in the integral on the right, we have, by a theorem of advanced calculus,

$$\int_{v_1}^{u_2}\int_{v_1}^{v_2} g(u, v)\,du\,dv = \int_{u_1}^{u_2}\int_{v_1}^{v_2} f[\psi_1(u, v), \psi_2(u, v)]J\,du\,dv$$

where

$$J = \frac{\partial(x, y)}{\partial(u, v)}$$

is the *Jacobian*. Thus

$$g(u, v) = f[\psi_1(u, v), \psi_2(u, v)]J$$

which is (36), page 42, in the case where $J > 0$. Similarly, we can prove (36) for the case where $J < 0$.

2.18. The probability function of a random variable X is

$$f(x) = \begin{cases} 2^{-x} & x = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Find the probability function for the random variable $U = X^4 + 1$.

Since $U = X^4 + 1$, the relationship between the values u and x of the random variables U and X is given by $u = x^4 + 1$ or $x = \sqrt[4]{u - 1}$, where $u = 2, 17, 82, \ldots$ and the real positive root is taken. Then the required probability function for U is given by

$$g(u) = \begin{cases} 2^{-\sqrt[4]{u-1}} & u = 2, 17, 82, \ldots \\ 0 & \text{otherwise} \end{cases}$$

using Theorem 2-1, page 42, or Problem 2.15.

2.19. The probability function of a random variable X is given by

$$f(x) = \begin{cases} x^2/81 & -3 < x < 6 \\ 0 & \text{otherwise} \end{cases}$$

Find the probability density for the random variable $U = \frac{1}{3}(12 - X)$.

We have $u = \frac{1}{3}(12 - x)$ or $x = 12 - 3u$. Thus to each value of x there is one and only one value of u and conversely. The values of u corresponding to $x = -3$ and $x = 6$ are $u = 5$ and $u = 2$, respectively. Since $\psi'(u) = dx/du = -3$, it follows by Theorem 2-3, page 42, or Problem 2.16 that the density function for U *is*

$$g(u) = \begin{cases} (12 - 3u)^2/27 & 2 < u < 5 \\ 0 & \text{otherwise} \end{cases}$$

Check: $$\int_2^5 \frac{(12 - 3u)^2}{27}\,du = -\frac{(12 - 3u)^3}{243}\bigg|_2^5 = 1$$

2.20. Find the probability density of the random variable $U = X^2$ where X is the random variable of Problem 2.19.

We have $u = x^2$ or $x = \pm\sqrt{u}$. Thus to each value of x there corresponds one and only one value of u, but to each value of $u \neq 0$ there correspond *two* values of x. The values of x for which $-3 < x < 6$ correspond to values of u for which $0 \le u < 36$ as shown in Fig. 2-12.

As seen in this figure, the interval $-3 < x \le 3$ corresponds to $0 \le u \le 9$ while $3 < x < 6$ corresponds to $9 < u < 36$. In this case we cannot use Theorem 2-3 directly but can proceed as follows. The distribution function for U is

$$G(u) = P(U \le u)$$

Now if $0 \le u \le 9$, we have

$$G(u) = P(U \le u) = P(X^2 \le u) = P(-\sqrt{u} \le X \le \sqrt{u})$$

$$= \int_{-\sqrt{u}}^{\sqrt{u}} f(x)\,dx$$

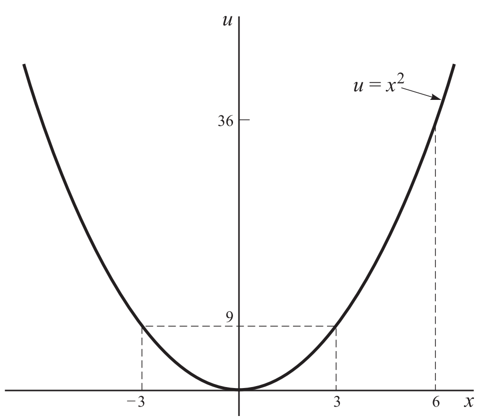


Fig. 2-12

But if $9 < u < 36$, we have

$$G(u) = P(U \le u) = P(-3 < X < \sqrt{u}) = \int_{-3}^{\sqrt{u}} f(x)\,dx$$

Since the density function $g(u)$ is the derivative of $G(u)$, we have, using (12),

$$g(u) = \begin{cases} \dfrac{f(\sqrt{u}) + f(-\sqrt{u})}{2\sqrt{u}} & 0 \le u \le 9 \\ \dfrac{f(\sqrt{u})}{2\sqrt{u}} & 9 < u < 36 \\ 0 & \text{otherwise} \end{cases}$$

Using the given definition of $f(x)$, this becomes

$$g(u) = \begin{cases} \sqrt{u}/81 & 0 \le u \le 9 \\ \sqrt{u}/162 & 9 < u < 36 \\ 0 & \text{otherwise} \end{cases}$$

Check:

$$\int_0^9 \frac{\sqrt{u}}{81}\,du + \int_9^{36} \frac{\sqrt{u}}{162}\,du = \left.\frac{2u^{3/2}}{243}\right|_0^9 + \left.\frac{u^{3/2}}{243}\right|_9^{36} = 1$$

2.21. If the random variables X and Y have joint density function

$$f(x, y) = \begin{cases} xy/96 & 0 < x < 4,\ 1 < y < 5 \\ 0 & \text{otherwise} \end{cases}$$

(see Problem 2.11), find the density function of $U = X + 2Y$.

Method 1

Let $u = x + 2y$, $v = x$, the second relation being chosen arbitrarily. Then simultaneous solution yields $x = v$, $y = \frac{1}{2}(u - v)$. Thus the region $0 < x < 4$, $1 < y < 5$ corresponds to the region $0 < v < 4$, $2 < u - v < 10$ shown shaded in Fig. 2-13.

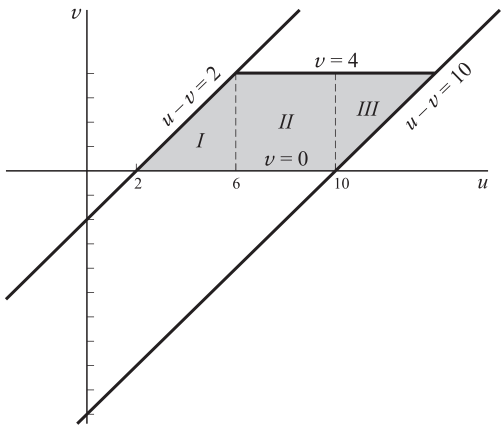


Fig. 2-13

The Jacobian is given by

$$J = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ \dfrac{1}{2} & -\dfrac{1}{2} \end{vmatrix} = -\frac{1}{2}$$

Then by Theorem 2-4 the joint density function of U and V is

$$g(u, v) = \begin{cases} v(u - v)/384 & 2 < u - v < 10, 0 < v < 4 \\ 0 & \text{otherwise} \end{cases}$$

The marginal density function of U is given by

$$g_1(u) = \begin{cases} \displaystyle\int_{v=0}^{u-2} \frac{v(u - v)}{384}\,dv & 2 < u < 6 \\ \displaystyle\int_{v=0}^{4} \frac{v(u - v)}{384}\,dv & 6 < u < 10 \\ \displaystyle\int_{v=u-10}^{4} \frac{v(u - v)}{384}\,dv & 10 < u < 14 \\ 0 & \text{otherwise} \end{cases}$$

as seen by referring to the shaded regions *I*, *II*, *III* of Fig. 2-13. Carrying out the integrations, we find

$$g_1(u) = \begin{cases} (u - 2)^2(u + 4)/2304 & 2 < u < 6 \\ (3u - 8)/144 & 6 < u < 10 \\ (348u - u^3 - 2128)/2304 & 10 < u < 14 \\ 0 & \text{otherwise} \end{cases}$$

A check can be achieved by showing that the integral of $g_1(u)$ is equal to 1.

Method 2
The distribution function of the random variable $X + 2Y$ is given by

$$P(X + 2Y \leq u) = \iint\limits_{x+2y \leq u} f(x, y)\,dx\,dy = \iint\limits_{\substack{x+2y \leq u \\ 0<x<4 \\ 1<y<5}} \frac{xy}{96}\,dx\,dy$$

For $2 < u < 6$, we see by referring to Fig. 2-14, that the last integral equals

$$\int_{x=0}^{u-2}\int_{y=1}^{(u-x)/2} \frac{xy}{96}\,dx\,dy = \int_{x=0}^{u-2}\left[\frac{x(u - x)^2}{768} - \frac{x}{192}\right]dx$$

The derivative of this with respect to u is found to be $(u - 2)^2(u + 4)/2304$. In a similar manner we can obtain the result of Method 1 for $6 < u < 10$, etc.

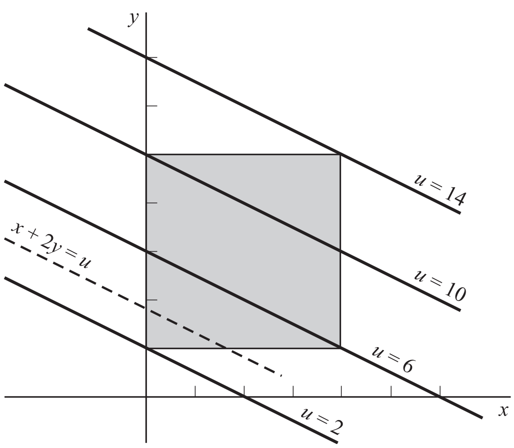


Fig. 2-14

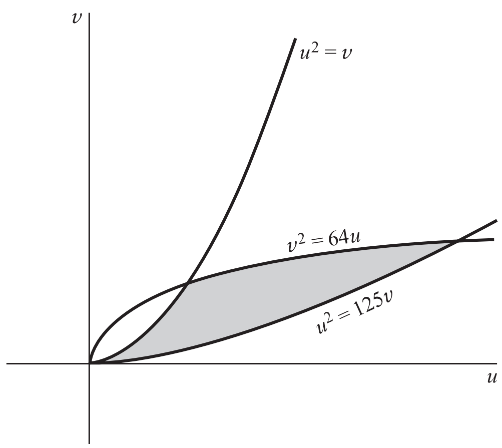


Fig. 2-15

2.22. If the random variables X and Y have joint density function

$$f(x, y) = \begin{cases} xy/96 & 0 < x < 4, 1 < y < 5 \\ 0 & \text{otherwise} \end{cases}$$

(see Problem 2.11), find the joint density function of $U = XY^2$, $V = X^2Y$.

Consider $u = xy^2$, $v = x^2y$. Dividing these equations, we obtain $y/x = u/v$ so that $y = ux/v$. This leads to the simultaneous solution $x = v^{2/3}u^{-1/3}$, $y = u^{2/3}v^{-1/3}$. The image of $0 < x < 4$, $1 < y < 5$ in the uv-plane is given by

$$0 < v^{2/3}u^{-1/3} < 4 \qquad 1 < u^{2/3}v^{-1/3} < 5$$

which are equivalent to

$$v^2 < 64u \qquad v < u^2 < 125v$$

This region is shown shaded in Fig. 2-15.

The Jacobian is given by

$$J = \begin{vmatrix} -\frac{1}{3}v^{2/3}u^{-4/3} & \frac{2}{3}v^{-1/3}u^{-1/3} \\ \frac{2}{3}u^{-1/3}v^{-1/3} & -\frac{1}{3}u^{2/3}v^{-4/3} \end{vmatrix} = -\frac{1}{3}u^{-2/3}v^{-2/3}$$

Thus the joint density function of U and V is, by Theorem 2-4,

$$g(u, v) = \begin{cases} \dfrac{(v^{2/3}u^{-1/3})(u^{2/3}v^{-1/3})}{96}\left(\tfrac{1}{3}u^{-2/3}v^{-2/3}\right) & v^2 < 64u, v < u^2 < 125v \\ 0 & \text{otherwise} \end{cases}$$

or

$$g(u, v) = \begin{cases} u^{-1/3}v^{-1/3}/288 & v^2 < 64u, \quad v < u^2 < 125v \\ 0 & \text{otherwise} \end{cases}$$

Convolutions

2.23. Let X and Y be random variables having joint density function $f(x, y)$. Prove that the density function of $U = X + Y$ is

$$g(u) = \int_{-\infty}^{\infty} f(v, u - v)\,dv$$

Method 1

Let $U = X + Y$, $V = X$, where we have arbitrarily added the second equation. Corresponding to these we have $u = x + y$, $v = x$ or $x = v$, $y = u - v$. The Jacobian of the transformation is given by

$$J = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1$$

Thus by Theorem 2-4, page 42, the joint density function of U and V is

$$g(u, v) = f(v, u - v)$$

It follows from (26), page 41, that the marginal density function of U is

$$g(u) = \int_{-\infty}^{\infty} f(v, u - v)\,dv$$

Method 2

The distribution function of $U = X + Y$ is equal to the double integral of $f(x, y)$ taken over the region defined by $x + y \leq u$, i.e.,

$$G(u) = \iint\limits_{x+y \leq u} f(x, y)\,dx\,dy$$

Since the region is below the line $x + y = u$, as indicated by the shading in Fig. 2-16, we see that

$$G(u) = \int_{x=-\infty}^{\infty} \left[\int_{y=-\infty}^{u-x} f(x, y)\,dy \right] dx$$

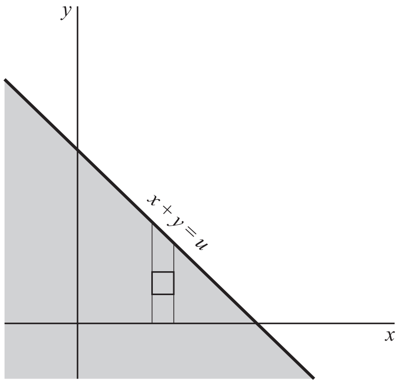


Fig. 2-16

The density function of U is the derivative of $G(u)$ with respect to u and is given by

$$g(u) = \int_{-\infty}^{\infty} f(x, u - x)\,dx$$

using (12) first on the x integral and then on the y integral.

2.24. Work Problem 2.23 if X and Y are independent random variables having density functions $f_1(x), f_2(y)$, respectively.

In this case the joint density function is $f(x, y) = f_1(x) f_2(y)$, so that by Problem 2.23 the density function of $U = X + Y$ is

$$g(u) = \int_{-\infty}^{\infty} f_1(v) f_2(u - v)\,dv = f_1 * f_2$$

which is the *convolution* of f_1 and f_2.

2.25. If X and Y are independent random variables having density functions

$$f_1(x) = \begin{cases} 2e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad f_2(y) = \begin{cases} 3e^{-3y} & y \geq 0 \\ 0 & y < 0 \end{cases}$$

find the density function of their sum, $U = X + Y$.

By Problem 2.24 the required density function is the convolution of f_1 and f_2 and is given by

$$g(u) = f_1 * f_2 = \int_{-\infty}^{\infty} f_1(v) f_2(u - v)\,dv$$

In the integrand f_1 vanishes when $v < 0$ and f_2 vanishes when $v > u$. Hence

$$\begin{aligned} g(u) &= \int_0^u (2e^{-2v})(3e^{-3(u-v)})\,dv \\ &= 6e^{-3u} \int_0^u e^v\,dv = 6e^{-3u}(e^u - 1) = 6(e^{-2u} - e^{3u}) \end{aligned}$$

if $u \geq 0$ and $g(u) = 0$ if $u < 0$.

Check: $$\int_{-\infty}^{\infty} g(u)\,du = 6\int_{0}^{\infty} (e^{-2u} - e^{-3u})\,du = 6\left(\frac{1}{2} - \frac{1}{3}\right) = 1$$

2.26. Prove that $f_1 * f_2 = f_2 * f_1$ (Property 1, page 43).

We have

$$f_1 * f_2 = \int_{v=-\infty}^{\infty} f_1(v) f_2(u - v)\,dv$$

Letting $w = u - v$ so that $v = u - w$, $dv = -dw$, we obtain

$$f_1 * f_2 = \int_{w=\infty}^{-\infty} f_1(u - w) f_2(w)(-dw) = \int_{w=-\infty}^{\infty} f_2(w) f_1(u - w)\,dw = f_2 * f_1$$

Conditional distributions

2.27. Find (a) $f(y\,|\,2)$, (b) $P(Y = 1\,|\,X = 2)$ for the distribution of Problem 2.8.

(a) Using the results in Problems 2.8 and 2.9, we have

$$f(y\,|\,x) = \frac{f(x, y)}{f_1(x)} = \frac{(2x + y)/42}{f_1(x)}$$

so that with $x = 2$

$$f(y\,|\,2) = \frac{(4 + y)/42}{11/21} = \frac{4 + y}{22}$$

(b) $$P(Y = 1\,|\,X = 2) = f(1\,|\,2) = \frac{5}{22}$$

2.28. If X and Y have the joint density function

$$f(x, y) = \begin{cases} \frac{3}{4} + xy & 0 < x < 1,\ 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

find (a) $f(y\,|\,x)$, (b) $P(Y > \frac{1}{2}\,|\,\frac{1}{2} < X < \frac{1}{2} + dx)$.

(a) For $0 < x < 1$,

$$f_1(x) = \int_0^1 \left(\frac{3}{4} + xy\right) dy = \frac{3}{4} + \frac{x}{2}$$

and $$f(y\,|\,x) = \frac{f(x, y)}{f_1(x)} = \begin{cases} \dfrac{3 + 4xy}{3 + 2x} & 0 < y < 1 \\ 0 & \text{other } y \end{cases}$$

For other values of x, $f(y\,|\,x)$ is not defined.

(b) $$P(Y > \tfrac{1}{2}\,|\,\tfrac{1}{2} < X < \tfrac{1}{2} + dx) = \int_{1/2}^{\infty} f(y\,|\,\tfrac{1}{2})\,dy = \int_{1/2}^{1} \frac{3 + 2y}{4}\,dy = \frac{9}{16}$$

2.29. The joint density function of the random variables X and Y is given by

$$f(x, y) = \begin{cases} 8xy & 0 \leq x \leq 1,\ 0 \leq y \leq x \\ 0 & \text{otherwise} \end{cases}$$

Find (a) the marginal density of X, (b) the marginal density of Y, (c) the conditional density of X, (d) the conditional density of Y.

The region over which $f(x, y)$ is different from zero is shown shaded in Fig. 2-17.

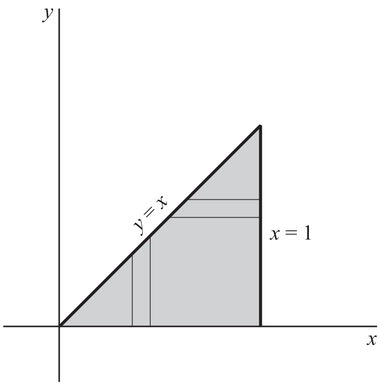


Fig. 2-17

(a) To obtain the marginal density of X, we fix x and integrate with respect to y from 0 to x as indicated by the vertical strip in Fig. 2-17. The result is

$$f_1(x) = \int_{y=0}^{x} 8xy\,dy = 4x^3$$

for $0 < x < 1$. For all other values of x, $f_1(x) = 0$.

(b) Similarly, the marginal density of Y is obtained by fixing y and integrating with respect to x from $x = y$ to $x = 1$, as indicated by the horizontal strip in Fig. 2-17. The result is, for $0 < y < 1$,

$$f_2(y) = \int_{x=y}^{1} 8xy\,dx = 4y(1 - y^2)$$

For all other values of y, $f_2(y) = 0$.

(c) The conditional density function of X is, for $0 < y < 1$,

$$f_1(x \mid y) = \frac{f(x, y)}{f_2(y)} = \begin{cases} 2x/(1 - y^2) & y \le x \le 1 \\ 0 & \text{other } x \end{cases}$$

The conditional density function is not defined when $f_2(y) = 0$.

(d) The conditional density function of Y is, for $0 < x < 1$,

$$f_2(y \mid x) = \frac{f(x, y)}{f_1(x)} = \begin{cases} 2y/x^2 & 0 \le y \le x \\ 0 & \text{other } y \end{cases}$$

The conditional density function is not defined when $f_1(x) = 0$.

Check:

$$\int_0^1 f_1(x)\,dx = \int_0^1 4x^3\,dx = 1, \quad \int_0^1 f_2(y)\,dy = \int_0^1 4y(1 - y^2)\,dy = 1$$

$$\int_y^1 f_1(x \mid y)\,dx = \int_y^1 \frac{2x}{1 - y^2}\,dx = 1$$

$$\int_0^x f_2(y \mid x)\,dy = \int_0^x \frac{2y}{x^2}\,dy = 1$$

2.30. Determine whether the random variables of Problem 2.29 are independent.

In the shaded region of Fig. 2-17, $f(x, y) = 8xy$, $f_1(x) = 4x^3$, $f_2(y) = 4y\,(1 - y^2)$. Hence $f(x, y) \ne f_1(x) f_2(y)$, and thus X and Y are dependent.

It should be noted that it does not follow from $f(x, y) = 8xy$ that $f(x, y)$ can be expressed as a function of x alone times a function of y alone. This is because the restriction $0 \le y \le x$ occurs. If this were replaced by some restriction on y not depending on x (as in Problem 2.21), such a conclusion would be valid.

Applications to geometric probability

2.31. A person playing darts finds that the probability of the dart striking between r and $r + dr$ is

$$P(r \le R \le r + dr) = c\left[1 - \left(\frac{r}{a}\right)^2\right]dr$$

Here, R is the distance of the hit from the center of the target, c is a constant, and a is the radius of the target (see Fig. 2-18). Find the probability of hitting the bull's-eye, which is assumed to have radius b. Assume that the target is always hit.

The density function is given by

$$f(r) = c\left[1 - \left(\frac{r}{a}\right)^2\right]$$

Since the target is always hit, we have

$$c\int_0^a \left[1 - \left(\frac{r}{a}\right)^2\right]dr = 1$$

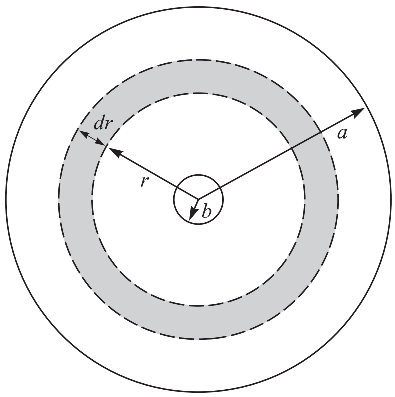


Fig. 2-18

from which $c = 3/2a$. Then the probability of hitting the bull's-eye is

$$\int_0^b f(r)\,dr = \frac{3}{2a}\int_0^b \left[1 - \left(\frac{r}{a}\right)^2\right]dr = \frac{b(3a^2 - b^2)}{2a^3}$$

2.32. Two points are selected at random in the interval $0 \le x \le 1$. Determine the probability that the sum of their squares is less than 1.

Let X and Y denote the random variables associated with the given points. Since equal intervals are assumed to have equal probabilities, the density functions of X and Y are given, respectively, by

(1) $$f_1(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \qquad f_2(y) = \begin{cases} 1 & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Then since X and Y are independent, the joint density function is given by

(2) $$f(x, y) = f_1(x)f_2(y) = \begin{cases} 1 & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

It follows that the required probability is given by

(3) $$P(X^2 + Y^2 \le 1) = \iint_{\mathcal{R}} dx\,dy$$

where $\mathcal{R}$ is the region defined by $x^2 + y^2 \le 1$, $x \ge 0$, $y \ge 0$, which is a quarter of a circle of radius 1 (Fig. 2-19). Now since (3) represents the area of $\mathcal{R}$, we see that the required probability is $\pi/4$.

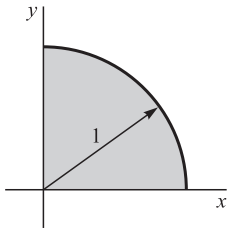


Fig. 2-19

Miscellaneous problems

2.33. Suppose that the random variables X and Y have a joint density function given by

$$f(x, y) = \begin{cases} c(2x + y) & 2 < x < 6, 0 < y < 5 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) the constant c, (b) the marginal distribution functions for X and Y, (c) the marginal density functions for X and Y, (d) $P(3 < X < 4, Y > 2)$, (e) $P(X > 3)$, (f) $P(X + Y > 4)$, (g) the joint distribution function, (h) whether X and Y are independent.

(a) The total probability is given by

$$\int_{x=2}^{6}\int_{y=0}^{5} c(2x + y)\,dx\,dy = \int_{x=2}^{6} c\left(2xy + \frac{y^2}{2}\right)\Big|_0^5 dx$$

$$= \int_{x=2}^{6} c\left(10x + \frac{25}{2}\right)dx = 210c$$

For this to equal 1, we must have $c = 1/210$.

(b) The marginal distribution function for X is

$$F_1(x) = P(X \leq x) = \int_{u=-\infty}^{x}\int_{v=-\infty}^{\infty} f(u, v)\,du\,dv$$

$$= \begin{cases} \displaystyle\int_{u=-\infty}^{x}\int_{v=-\infty}^{\infty} 0\,du\,dv = 0 & x < 2 \\ \displaystyle\int_{u=2}^{x}\int_{v=0}^{5} \frac{2u + v}{210}\,du\,dv = \frac{2x^2 + 5x - 18}{84} & 2 \leq x < 6 \\ \displaystyle\int_{u=2}^{6}\int_{v=0}^{5} \frac{2u + v}{210}\,du\,dv = 1 & x \geq 6 \end{cases}$$

The marginal distribution function for Y is

$$F_2(y) = P(Y \leq y) = \int_{u=-\infty}^{\infty}\int_{v=-\infty}^{y} f(u, v)\,du\,dv$$

$$= \begin{cases} \displaystyle\int_{u=-\infty}^{\infty}\int_{v=-8}^{y} 0\,du\,dv = 0 & y < 0 \\ \displaystyle\int_{u=0}^{6}\int_{v=0}^{y} \frac{2u + v}{210}\,du\,dv = \frac{y^2 + 16y}{105} & 0 \leq y < 5 \\ \displaystyle\int_{u=2}^{6}\int_{v=0}^{5} \frac{2u + v}{210}\,du\,dv = 1 & y \geq 5 \end{cases}$$

(c) The marginal density function for X is, from part (b),

$$f_1(x) = \frac{d}{dx}F_1(x) = \begin{cases} (4x + 5)/84 & 2 < x < 6 \\ 0 & \text{otherwise} \end{cases}$$

The marginal density function for Y is, from part (b),

$$f_2(y) = \frac{d}{dy}F_2(y) = \begin{cases} (2y + 16)/105 & 0 < y < 5 \\ 0 & \text{otherwise} \end{cases}$$

(d) $$P(3 < X < 4, Y > 2) = \frac{1}{210}\int_{x=3}^{4}\int_{y=2}^{5}(2x + y)\,dx\,dy = \frac{3}{20}$$

(e) $$P(X > 3) = \frac{1}{210}\int_{x=3}^{6}\int_{y=0}^{5}(2x + y)\,dx\,dy = \frac{23}{28}$$

(f) $$P(X + Y > 4) = \iint_{\mathscr{R}} f(x, y)\,dx\,dy$$

where $\mathscr{R}$ is the shaded region of Fig. 2-20. Although this can be found, it is easier to use the fact that

$$P(X + Y > 4) = 1 - P(X + Y \le 4) = 1 - \iint_{\mathscr{R}} f(x, y)\,dx\,dy$$

where $\mathscr{R}'$ is the cross-hatched region of Fig. 2-20. We have

$$P(X + Y \le 4) = \frac{1}{210}\int_{x=2}^{4}\int_{y=0}^{4-x}(2x + y)\,dx\,dy = \frac{2}{35}$$

Thus $P(X + Y > 4) = 33/35$.

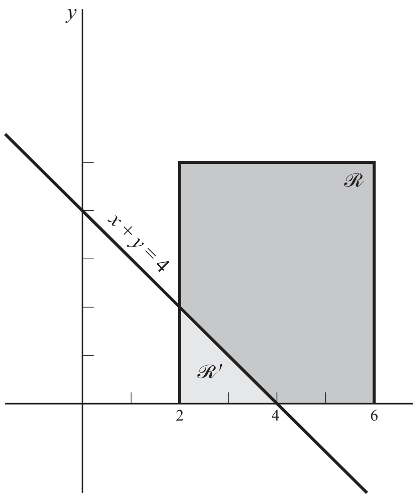


Fig. 2-20

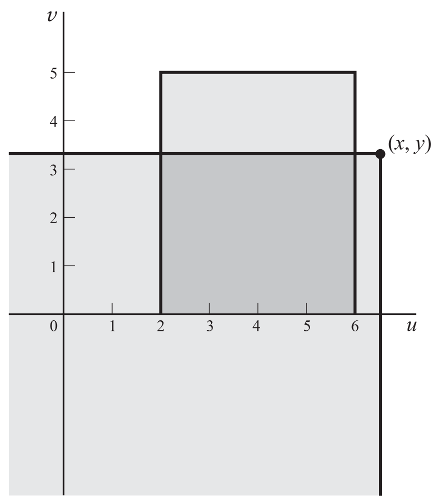


Fig. 2-21

(g) The joint distribution function is

$$F(x, y) = P(X \le x, Y \le y) = \int_{u=-\infty}^{x}\int_{v=-\infty}^{y} f(u, v)\,du\,dv$$

In the uv plane (Fig. 2-21) the region of integration is the intersection of the quarter plane $u \le x$, $v \le y$ and the rectangle $2 < u < 6$, $0 < v < 5$ [over which $f(u, v)$ is nonzero]. For (x, y) located as in the figure, we have

$$F(x, y) = \int_{u=2}^{6}\int_{v=0}^{y}\frac{2u + v}{210}\,du\,dv = \frac{16y + y^2}{105}$$

When (x, y) lies inside the rectangle, we obtain another expression, etc. The complete results are shown in Fig. 2-22.

(h) The random variables are dependent since

$$f(x, y) \neq f_1(x)f_2(y)$$

or equivalently, $F(x, y) \neq F_1(x)F_2(y)$.

2.34. Let X have the density function

$$f(x) = \begin{cases} 6x(1-x) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find a function $Y = h(X)$ which has the density function

$$g(y) = \begin{cases} 12y^3(1-y^2) & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

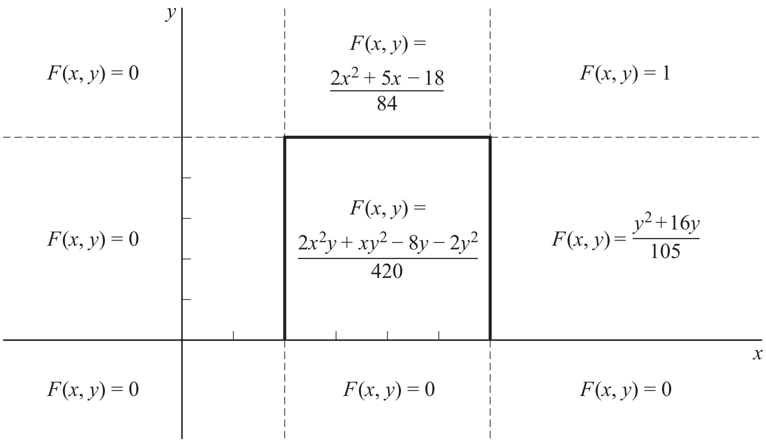


Fig. 2-22

We assume that the unknown function h is such that the intervals $X \leq x$ and $Y \leq y + h(x)$ correspond in a one-one, continuous fashion. Then $P(X \leq x) = P(Y \leq y)$, i.e., the distribution functions of X and Y must be equal. Thus, for $0 < x, y < 1$,

$$\int_0^x 6u(1-u)\,du = \int_0^y 12v^3(1-v^2)\,dv$$

or

$$3x^2 - 2x^3 = 3y^4 - 2y^6$$

By inspection, $x = y^2$ or $y = h(x) = +\sqrt{x}$ is a solution, and this solution has the desired properties. Thus $Y = +\sqrt{X}$.

2.35. Find the density function of $U = XY$ if the joint density function of X and Y is $f(x, y)$.

Method 1

Let $U = XY$ and $V = X$, corresponding to which $u = xy$, $v = x$ or $x = v$, $y = u/v$. Then the Jacobian is given by

$$J = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ v^{-1} & -uv^{-2} \end{vmatrix} = -v^{-1}$$

Thus the joint density function of U and V is

$$g(u, v) = \frac{1}{|v|} f\left(v, \frac{u}{v}\right)$$

from which the marginal density function of U is obtained as

$$g(u) = \int_{-\infty}^{\infty} g(u, v)\,dv = \int_{-\infty}^{\infty} \frac{1}{|v|} f\left(v, \frac{u}{v}\right) dv$$

Method 2

The distribution function of U is

$$G(u) = \iint_{xy \le u} f(x, y)\,dx\,dy$$

For $u \geq 0$, the region of integration is shown shaded in Fig. 2-23. We see that

$$G(u) = \int_{-\infty}^{0} \left[\int_{u/x}^{\infty} f(x, y)\,dy \right] dx + \int_{0}^{\infty} \left[\int_{-\infty}^{u/x} f(x, y)\,dy \right] dx$$

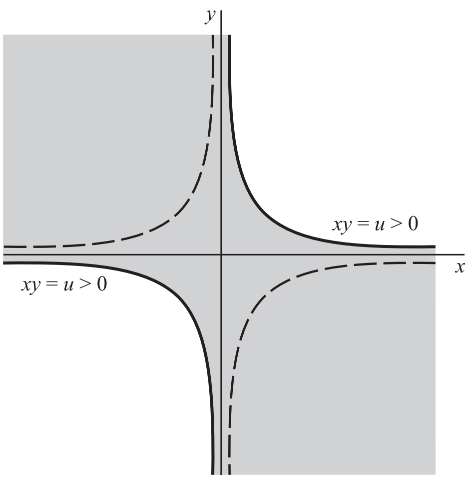


Fig. 2-23

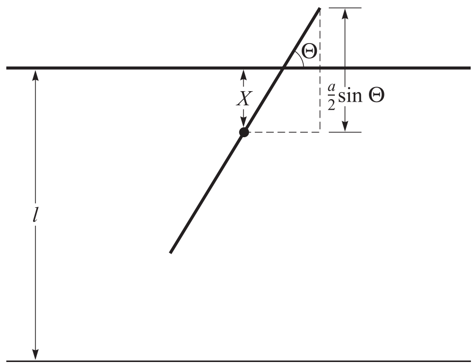


Fig. 2-24

Differentiating with respect to u, we obtain

$$g(u) = \int_{-\infty}^{0} \left(\frac{-1}{x}\right) f\left(x, \frac{u}{x}\right) dx + \int_{0}^{\infty} \frac{1}{x} f\left(x, \frac{u}{x}\right) dx = \int_{-\infty}^{\infty} \frac{1}{|x|} f\left(x, \frac{u}{x}\right) dx$$

The same result is obtained for $u < 0$, when the region of integration is bounded by the dashed hyperbola in Fig. 2-24.

2.36. A floor has parallel lines on it at equal distances l from each other. A needle of length $a < l$ is dropped at random onto the floor. Find the probability that the needle will intersect a line. (This problem is known as *Buffon's needle problem.*)

Let X be a random variable that gives the distance of the midpoint of the needle to the nearest line (Fig. 2-24). Let Θ be a random variable that gives the acute angle between the needle (or its extension) and the line. We denote by x and θ any particular values of X and Θ. It is seen that X can take on any value between 0 and $l/2$, so that $0 \leq x \leq l/2$. Also Θ can take on any value between 0 and $\pi/2$. It follows that

$$P(x < X \leq x + dx) = \frac{2}{l}\,dx \qquad P(\theta \leq \Theta + d\theta) = \frac{2}{\pi}\,d\theta$$

i.e., the density functions of X and Θ are given by $f_1(x) = 2/l$, $f_2(\theta) = 2/\pi$. As a check, we note that

$$\int_{0}^{l/2} \frac{2}{l}\,dx = 1 \qquad \int_{0}^{\pi/2} \frac{2}{\pi}\,d\theta = 1$$

Since X and Θ are independent the joint density function is

$$f(x, \theta) = \frac{2}{l} \cdot \frac{2}{\pi} = \frac{4}{l\pi}$$

From Fig. 2-24 it is seen that the needle actually hits a line when $X \leq (a/2) \sin \Theta$. The probability of this event is given by

$$\frac{4}{l\pi} \int_{\theta=0}^{\pi/2} \int_{x=0}^{(a/2)\sin\theta} dx\, d\theta = \frac{2a}{l\pi}$$

When the above expression is equated to the frequency of hits observed in actual experiments, accurate values of π are obtained. This indicates that the probability model described above is appropriate.

2.37. Two people agree to meet between 2:00 P.M. and 3:00 P.M., with the understanding that each will wait no longer than 15 minutes for the other. What is the probability that they will meet?

Let X and Y be random variables representing the times of arrival, measured in fractions of an hour after 2:00 P.M., of the two people. Assuming that equal intervals of time have equal probabilities of arrival, the density functions of X and Y are given respectively by

$$f_1(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_2(y) = \begin{cases} 1 & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Then, since X and Y are independent, the joint density function is

(1) $$f(x, y) = f_1(x) f_2(y) = \begin{cases} 1 & 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Since 15 minutes $= \frac{1}{4}$ hour, the required probability is

(2) $$P\left(|X - Y| \leq \frac{1}{4}\right) = \iint_{\mathscr{R}} dx\, dy$$

where 5 is the region shown shaded in Fig. 2-25. The right side of (2) is the area of this region, which is equal to $1 - (\frac{3}{4})(\frac{3}{4}) = \frac{7}{16}$, since the square has area 1, while the two corner triangles have areas $\frac{1}{2}(\frac{3}{4})(\frac{3}{4})$ each. Thus the required probability is $7/16$.

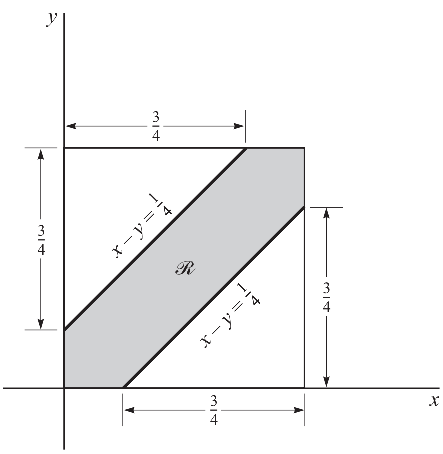


Fig. 2-25

SUPPLEMENTARY PROBLEMS

Discrete random variables and probability distributions

2.38. A coin is tossed three times. If X is a random variable giving the number of heads that arise, construct a table showing the probability distribution of X.

2.39. An urn holds 5 white and 3 black marbles. If 2 marbles are to be drawn at random without replacement and X denotes the number of white marbles, find the probability distribution for X.

2.40. Work Problem 2.39 if the marbles are to be drawn with replacement.

2.41. Let Z be a random variable giving the number of heads minus the number of tails in 2 tosses of a fair coin. Find the probability distribution of Z. Compare with the results of Examples 2.1 and 2.2.

2.42. Let X be a random variable giving the number of aces in a random draw of 4 cards from an ordinary deck of 52 cards. Construct a table showing the probability distribution of X.

Discrete distribution functions

2.43. The probability function of a random variable X is shown in Table 2-7. Construct a table giving the distribution function of X.

Table 2-7

x	1	2	3
$f(x)$	1/2	1/3	1/6

Table 2-8

x	1	2	3	4
$F(x)$	1/8	3/8	3/4	1

2.44. Obtain the distribution function for (a) Problem 2.38, (b) Problem 2.39, (c) Problem 2.40.

2.45. Obtain the distribution function for (a) Problem 2.41, (b) Problem 2.42.

2.46. Table 2-8 shows the distribution function of a random variable X. Determine (a) the probability function, (b) $P(1 \le X \le 3)$, (c) $P(X \ge 2)$, (d) $P(X < 3)$, (e) $P(X > 1.4)$.

Continuous random variables and probability distributions

2.47. A random variable X has density function

$$f(x) = \begin{cases} ce^{-3x} & x > 0 \\ 0 & x \le 0 \end{cases}$$

Find (a) the constant c, (b) $P(1 < X < 2)$, (c) $P(X \ge 3)$, (d) $P(X < 1)$.

2.48. Find the distribution function for the random variable of Problem 2.47. Graph the density and distribution functions, describing the relationship between them.

2.49. A random variable X has density function

$$f(x) = \begin{cases} cx^2 & 1 \le x \le 2 \\ cx & 2 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) the constant c, (b) $P(X > 2)$, (c) $P(1/2 < X < 3/2)$.

2.50. Find the distribution function for the random variable X of Problem 2.49.

2.51. The distribution function of a random variable X is given by

$$F(x) = \begin{cases} cx^3 & 0 \le x < 3 \\ 1 & x \ge 3 \\ 0 & x < 0 \end{cases}$$

If $P(X = 3) = 0$, find (a) the constant c, (b) the density function, (c) $P(X > 1)$, (d) $P(1 < X < 2)$.

2.52. Can the function

$$F(x) = \begin{cases} c(1 - x^2) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

be a distribution function? Explain.

2.53. Let X be a random variable having density function

$$f(x) = \begin{cases} cx & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) the value of the constant c, (b) $P(\frac{1}{2} < X < \frac{3}{2})$, (c) $P(X > 1)$, (d) the distribution function.

Joint distributions and independent variables

2.54. The joint probability function of two discrete random variables X and Y is given by $f(x, y) = cxy$ for $x = 1, 2, 3$ and $y = 1, 2, 3$, and equals zero otherwise. Find (a) the constant c, (b) $P(X = 2, Y = 3)$, (c) $P(1 \le X \le 2, Y \le 2)$, (d) $P(X \ge 2)$, (e) $P(Y < 2)$, (f) $P(X = 1)$, (g) $P(Y = 3)$.

2.55. Find the marginal probability functions of (a) X and (b) Y for the random variables of Problem 2.54. (c) Determine whether X and Y are independent.

2.56. Let X and Y be continuous random variables having joint density function

$$f(x, y) = \begin{cases} c(x^2 + y^2) & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Determine (a) the constant c, (b) $P(X < \frac{1}{2}, Y > \frac{1}{2})$, (c) $P(\frac{1}{4} < X < \frac{3}{4})$, (d) $P(Y < \frac{1}{2})$, (e) whether X and Y are independent.

2.57. Find the marginal distribution functions (a) of X and (b) of Y for the density function of Problem 2.56.

Conditional distributions and density functions

2.58. Find the conditional probability function (a) of X given Y, (b) of Y given X, for the distribution of Problem 2.54.

2.59. Let

$$f(x, y) = \begin{cases} x + y & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the conditional density function of (a) X given Y, (b) Y given X.

2.60. Find the conditional density of (a) X given Y, (b) Y given X, for the distribution of Problem 2.56.

2.61. Let

$$f(x, y) = \begin{cases} e^{-(x+y)} & x \ge 0, y \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

be the joint density function of X and Y. Find the conditional density function of (a) X given Y, (b) Y given X.

Change of variables

2.62. Let X have density function

$$f(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Find the density function of $Y = X^2$.

2.63. (a) If the density function of X is $f(x)$ find the density function of X^3. (b) Illustrate the result in part (a) by choosing

$$f(x) = \begin{cases} 2e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

and check the answer.

2.64. If X has density function $f(x) = 2(\pi)^{-1/2}e^{-x^2/2}$, $-\infty < x < \infty$, find the density function of $Y = X^2$.

2.65. Verify that the integral of $g_1(u)$ in Method 1 of Problem 2.21 is equal to 1.

2.66. If the density of X is $f(x) = 1/\pi(x^2 + 1)$, $-\infty < x < \infty$, find the density of $Y = \tan^{-1} X$.

2.67. Complete the work needed to find $g_1(u)$ in Method 2 of Problem 2.21 and check your answer.

2.68. Let the density of X be

$$f(x) = \begin{cases} 1/2 & -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the density of (a) $3X - 2$, (b) $X^3 + 1$.

2.69. Check by direct integration the joint density function found in Problem 2.22.

2.70. Let X and Y have joint density function

$$f(x, y) = \begin{cases} e^{-(x+y)} & x \geq 0, y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

If $U = X/Y$, $V = X + Y$, find the joint density function of U and V.

2.71. Use Problem 2.22 to find the density function of (a) $U = XY^2$, (b) $V = X^2Y$.

2.72. Let X and Y be random variables having joint density function $f(x, y) = (2\pi)^{-1}e^{-(x^2+y^2)}$, $-\infty < x < \infty$, $-\infty < y < \infty$. If R and Θ are new random variables such that $X = R\cos\Theta$, $Y = R\sin\Theta$, show that the density function of R is

$$g(r) = \begin{cases} re^{-r^2/2} & r \geq 0 \\ 0 & r < 0 \end{cases}$$

2.73. Let
$$f(x, y) = \begin{cases} 1 & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$
be the joint density function of X and Y. Find the density function of $Z = XY$.

Convolutions

2.74. Let X and Y be identically distributed independent random variables with density function
$$f(t) = \begin{cases} 1 & 0 \le t \le 1 \\ 0 & \text{otherwise} \end{cases}$$
Find the density function of $X + Y$ and check your answer.

2.75. Let X and Y be identically distributed independent random variables with density function
$$f(t) = \begin{cases} e^{-t} & t \ge 0 \\ 0 & \text{otherwise} \end{cases}$$
Find the density function of $X + Y$ and check your answer.

2.76. Work Problem 2.21 by first making the transformation $2Y = Z$ and then using convolutions to find the density function of $U = X + Z$.

2.77. If the independent random variables X_1 and X_2 are identically distributed with density function
$$f(t) = \begin{cases} te^{-t} & t \ge 0 \\ 0 & t < 0 \end{cases}$$
find the density function of $X_1 + X_2$.

Applications to geometric probability

2.78. Two points are to be chosen at random on a line segment whose length is $a > 0$. Find the probability that the three line segments thus formed will be the sides of a triangle.

2.79. It is known that a bus will arrive at random at a certain location sometime between 3:00 P.M. and 3:30 P.M. A man decides that he will go at random to this location between these two times and will wait at most 5 minutes for the bus. If he misses it, he will take the subway. What is the probability that he will take the subway?

2.80. Two line segments, AB and CD, have lengths 8 and 6 units, respectively. Two points P and Q are to be chosen at random on AB and CD, respectively. Show that the probability that the area of a triangle will have height AP and that the base CQ will be greater than 12 square units is equal to $(1 - \ln 2)/2$.

Miscellaneous problems

2.81. Suppose that $f(x) = c/3^x$, $x = 1, 2, \ldots$, is the probability function for a random variable X. (a) Determine c. (b) Find the distribution function. (c) Graph the probability function and the distribution function. (d) Find $P(2 \le X < 5)$. (e) Find $P(X \ge 3)$.

2.82. Suppose that
$$f(x) = \begin{cases} cxe^{-2x} & x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$
is the density function for a random variable X. (a) Determine c. (b) Find the distribution function. (c) Graph the density function and the distribution function. (d) Find $P(X \ge 1)$. (e) Find $P(2 \le X < 3)$.

2.83. The probability function of a random variable X is given by

$$f(x) = \begin{cases} 2p & x = 1 \\ p & x = 2 \\ 4p & x = 3 \\ 0 & \text{otherwise} \end{cases}$$

where p is a constant. Find (a) $P(0 \le X < 3)$, (b) $P(X > 1)$.

2.84. (a) Prove that for a suitable constant c,

$$F(x) = \begin{cases} 0 & x \le 0 \\ c(1 - e^{-x})^2 & x > 0 \end{cases}$$

is the distribution function for a random variable X, and find this c. (b) Determine $P(1 < X < 2)$.

2.85. A random variable X has density function

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the density function of the random variable $Y = X^2$ and check your answer.

2.86. Two independent random variables, X and Y, have respective density functions

$$f(x) = \begin{cases} c_1 e^{-2x} & x > 0 \\ 0 & x \le 0 \end{cases} \qquad g(y) = \begin{cases} c_2 y e^{-3y} & y > 0 \\ 0 & y \le 0 \end{cases}$$

Find (a) c_1 and c_2, (b) $P(X + Y > 1)$, (c) $P(1 < X < 2, Y \ge 1)$, (d) $P(1 < X < 2)$, (e) $P(Y \ge 1)$.

2.87. In Problem 2.86 what is the relationship between the answers to (c), (d), and (e)? Justify your answer.

2.88. Let X and Y be random variables having joint density function

$$f(x, y) = \begin{cases} c(2x + y) & 0 < x < 1, 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) the constant c, (b) $P(X > \frac{1}{2}, Y < \frac{3}{2})$, (c) the (marginal) density function of X, (d) the (marginal) density function of Y.

2.89. In Problem 2.88 is $P(X > \frac{1}{2}, Y < \frac{3}{2}) = P(X > \frac{1}{2})P(Y < \frac{3}{2})$? Why?

2.90. In Problem 2.86 find the density function (a) of X^2, (b) of $X + Y$.

2.91. Let X and Y have joint density function

$$f(x, y) = \begin{cases} 1/y & 0 < x < y, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Determine whether X and Y are independent, (b) Find $P(X > \frac{1}{2})$. (c) Find $P(X < \frac{1}{2}, Y > \frac{1}{3})$. (d) Find $P(X + Y > \frac{1}{2})$.

2.92. Generalize (a) Problem 2.74 and (b) Problem 2.75 to three or more variables.

2.93. Let X and Y be identically distributed independent random variables having density function $f(u) = (2\pi)^{-1/2}e^{-u^2/2}$, $-\infty < u < \infty$. Find the density function of $Z = X^2 + Y^2$.

2.94. The joint probability function for the random variables X and Y is given in Table 2-9. (a) Find the marginal probability functions of X and Y. (b) Find $P(1 \leq X < 3, Y \geq 1)$. (c) Determine whether X and Y are independent.

Table 2-9

X \ Y	0	1	2
0	1/18	1/9	1/6
1	1/9	1/18	1/9
2	1/6	1/6	1/18

2.95. Suppose that the joint probability function of random variables X and Y is given by

$$f(x, y) = \begin{cases} cxy & 0 \leq x \leq 2, 0 \leq y \leq x \\ 0 & \text{otherwise} \end{cases}$$

(a) Determine whether X and Y are independent. (b) Find $P(\frac{1}{2} < X < 1)$. (c) Find $P(Y \geq 1)$. (d) Find $P(\frac{1}{2} < X < 1, Y \geq 1)$.

2.96. Let X and Y be independent random variables each having density function

$$f(u) = \frac{\lambda^u e^{-\lambda}}{u} \qquad u = 0, 1, 2, \ldots$$

where $\lambda > 0$. Prove that the density function of $X + Y$ is

$$g(u) = \frac{(2\lambda)^u e^{-2\lambda}}{u!} \qquad u = 0, 1, 2, \ldots$$

2.97. A stick of length L is to be broken into two parts. What is the probability that one part will have a length of more than double the other? State clearly what assumptions would you have made. Discuss whether you believe these assumptions are realistic and how you might improve them if they are not.

2.98. A floor is made up of squares of side l. A needle of length $a < l$ is to be tossed onto the floor. Prove that the probability of the needle intersecting at least one side is equal to $a(4l - a)/\pi l^2$.

2.99. For a needle of given length, what should be the side of a square in Problem 2.98 so that the probability of intersection is a maximum? Explain your answer.

2.100. Let
$$f(x, y, z) = \begin{cases} 24xy^2z^3 & 0 < x < 1, 0 < y < 1, 0 < z < 1 \\ 0 & \text{otherwise} \end{cases}$$

be the joint density function of three random variables X, Y, and Z. Find (a) $P(X > \frac{1}{2}, Y < \frac{1}{2}, Z > \frac{1}{2})$, (b) $P(Z < X + Y)$.

2.101. A cylindrical stream of particles, of radius a, is directed toward a hemispherical target ABC with center at O as indicated in Fig. 2-26. Assume that the distribution of particles is given by

$$f(r) = \begin{cases} 1/a & 0 < r < a \\ 0 & \text{otherwise} \end{cases}$$

where r is the distance from the axis OB. Show that the distribution of particles along the target is given by

$$g(\theta) = \begin{cases} \cos\theta & 0 < \theta < \pi/2 \\ 0 & \text{otherwise} \end{cases}$$

where θ is the angle that line OP (from O to any point P on the target) makes with the axis.

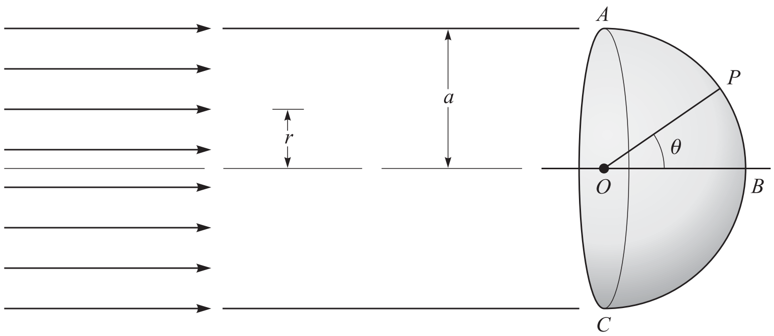


Fig. 2-26

2.102. In Problem 2.101 find the probability that a particle will hit the target between $\theta = 0$ and $\theta = \pi/4$.

2.103. Suppose that random variables X, Y, and Z have joint density function

$$f(x, y, z) = \begin{cases} 1 - \cos\pi x \cos\pi y \cos\pi z & 0 < x < 1, 0 < y < 1, 0 < z < 1 \\ 0 & \text{otherwise} \end{cases}$$

Show that although any two of these random variables are independent, i.e., their marginal density function factors, all three are not independent.

ANSWERS TO SUPPLEMENTARY PROBLEMS

2.38.

x	0	1	2	3
$f(x)$	1/8	3/8	3/8	1/8

2.39.

x	0	1	2
$f(x)$	3/28	15/28	5/14

2.40.

x	0	1	2
$f(x)$	9/64	15/32	25/64

2.42.

x	0	1	2	3	4
$f(x)$	$\frac{194{,}580}{270{,}725}$	$\frac{69{,}184}{270{,}725}$	$\frac{6768}{270{,}725}$	$\frac{192}{270{,}725}$	$\frac{1}{270{,}725}$

2.43.

x	0	1	2	3
$f(x)$	1/8	1/2	7/8	1

2.46. (a)

x	1	2	3	4
$f(x)$	1/8	1/4	3/8	1/4

(b) 3/4 (c) 7/8 (d) 3/8 (e) 7/8

2.47. (a) 3 (b) $e^{-3} - e^{-6}$ (c) e^{-9} (d) $1 - e^{-3}$ **2.48.** $F(x) = \begin{cases} 1 - e^{-3x} & x \geq 0 \\ 0 & x \leq 0 \end{cases}$

2.49. (a) 6/29 (b) 15/29 (c) 19/116 **2.50.** $F(x) = \begin{cases} 0 & x \leq 1 \\ (2x^3 - 2)/29 & 1 \leq x \leq 2 \\ (3x^2 + 2)/29 & 2 \leq x \leq 3 \\ 1 & x \geq 3 \end{cases}$

2.51. (a) 1/27 (b) $f(x) = \begin{cases} x^2/9 & 0 \leq x < 3 \\ 0 & \text{otherwise} \end{cases}$ (c) 26/27 (d) 7/27

2.53. (a) 1/2 (b) 1/2 (c) 3/4 (d) $F(x) = \begin{cases} 0 & x \leq 0 \\ x^2/4 & 0 \leq x \leq 2 \\ 1 & x \geq 2 \end{cases}$

2.54. (a) 1/36 (b) 1/6 (c) 1/4 (d) 5/6 (e) 1/6 (f) 1/6 (g) 1/2

2.55. (a) $f_1(x) = \begin{cases} x/6 & x = 1, 2, 3 \\ 0 & \text{other } x \end{cases}$ (b) $f_2(y) = \begin{cases} y/6 & y = 1, 2, 3 \\ 0 & \text{other } y \end{cases}$

2.56. (a) 3/2 (b) 1/4 (c) 29/64 (d) 5/16

2.57. (a) $F_1(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{2}(x^3 + x) & 0 \leq x \leq 1 \\ 1 & x \geq 1 \end{cases}$ (b) $F_2(y) = \begin{cases} 0 & y \leq 0 \\ \frac{1}{2}(y^3 + y) & 0 \leq y \leq 1 \\ 1 & y \geq 1 \end{cases}$

2.58. (a) $f(x \mid y) = f_1(x)$ for $y = 1, 2, 3$ (see Problem 2.55)

(b) $f(y \mid x) = f_2(y)$ for $x = 1, 2, 3$ (see Problem 2.55)

2.59. (a) $f(x|y) = \begin{cases} (x + y)/(y + \frac{1}{2}) & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{other } x, 0 \leq y \leq 1 \end{cases}$

(b) $f(y|x) = \begin{cases} (x + y)/(x + \frac{1}{2}) & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & 0 \leq x \leq 1, \text{ other } y \end{cases}$

2.60. (a) $f(x|y) = \begin{cases} (x^2 + y^2)/(y^2 + \frac{1}{3}) & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{other } x, 0 \leq y \leq 1 \end{cases}$

(b) $f(y|x) = \begin{cases} (x^2 + y^2)/(x^2 + \frac{1}{3}) & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & 0 \leq x \leq 1, \text{ other } y \end{cases}$

2.61. (a) $f(x \mid y) = \begin{cases} e^{-x} & x \geq 0, y \geq 0 \\ 0 & x < 0, y \geq 0 \end{cases}$ (b) $f(y \mid x) = \begin{cases} e^{-y} & x \geq 0, y \geq 0 \\ 0 & x \geq 0, y < 0 \end{cases}$

2.62. $e^{-\sqrt{y}}/2\sqrt{y}$ for $y > 0$; 0 otherwise **2.64.** $(2\pi)^{-1/2}y^{-1/2}e^{-y/2}$ for $y > 0$; 0 otherwise

2.66. $1/\pi$ for $-\pi/2 < y < \pi/2$; 0 otherwise

2.68. (a) $g(y) = \begin{cases} \frac{1}{6} & -5 < y < 1 \\ 0 & \text{otherwise} \end{cases}$ (b) $g(y) = \begin{cases} \frac{1}{6}(1 - y)^{-2/3} & 0 < y < 1 \\ \frac{1}{6}(y - 1)^{-2/3} & 1 < y < 2 \\ 0 & \text{otherwise} \end{cases}$

2.70. $ve^{-v}/(1 + u)^2$ for $u \geq 0$, $v \geq 0$; 0 otherwise

2.73. $g(z) = \begin{cases} -\ln z & 0 < z < 1 \\ 0 & \text{otherwise} \end{cases}$ **2.77.** $g(x) = \begin{cases} x^3 e^{-x}/6 & x \geq 0 \\ 0 & x < 0 \end{cases}$

2.74. $g(u) = \begin{cases} u & 0 \leq u \leq 1 \\ 2 - u & 1 \leq u \leq 2 \\ 0 & \text{otherwise} \end{cases}$ **2.78.** 1/4

2.75. $g(u) = \begin{cases} ue^{-u} & u \geq 0 \\ 0 & u < 0 \end{cases}$ **2.79.** 61/72

2.81. (a) 2 (b) $F(x) = \begin{cases} 0 & x < 1 \\ 1 - 3^{-y} & y \leq x < y + 1;\ y = 1, 2, 3, \ldots \end{cases}$ (d) 26/81 (e) 1/9

2.82. (a) 4 (b) $F(x) = \begin{cases} 1 - e^{-2x}(2x + 1) & x \geq 0 \\ 0 & x < 0 \end{cases}$ (d) $3e^{-2}$ (e) $5e^{-4} - 7e^{-6}$

2.83. (a) 3/7 (b) 5/7 **2.84.** (a) $c = 1$ (b) $e^{-4} - 3e^{-2} + 2e^{-1}$

2.86. (a) $c_1 = 2, c_2 = 9$ (b) $9e^{-2} - 14e^{-3}$ (c) $4e^{-5} - 4e^{-7}$ (d) $e^{-2} - e^{-4}$ (e) $4e^{-3}$

2.88. (a) 1/4 (b) 27/64 (c) $f_1(x) = \begin{cases} x + \frac{1}{2} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$ (d) $f_2(y) = \begin{cases} \frac{1}{4}(y + 1) & 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}$

2.90. (a) $\begin{cases} e^{-2y}/\sqrt{y} & y > 0 \\ 0 & \text{otherwise} \end{cases}$ (b) $\begin{cases} 18e^{-2u} & u > 0 \\ 0 & \text{otherwise} \end{cases}$

2.91. (b) $\frac{1}{2}(1 - \ln 2)$ (c) $\frac{1}{6} + \frac{1}{2}\ln 2$ (d) $\frac{1}{2}\ln 2$ **2.95.** (b) 15/256 (c) 9/16 (d) 0

2.93. $g(z) = \begin{cases} \frac{1}{2}e^{-z/2} & z \geq 0 \\ 0 & z < 0 \end{cases}$ **2.100.** (a) 45/512 (b) 1/14

2.94. (b) 7/18 **2.102.** $\sqrt{2}/2$

CHAPTER 3

Mathematical Expectation

Definition of Mathematical Expectation

A very important concept in probability and statistics is that of the *mathematical expectation*, *expected value*, or briefly the *expectation*, of a random variable. For a discrete random variable X having the possible values $x_1, \ldots, x_n$, the expectation of X is defined as

$$E(X) = x_1 P(X = x_1) + \cdots + x_n P(X = x_n) = \sum_{j=1}^{n} x_j P(X = x_j) \tag{1}$$

or equivalently, if $P(X = x_j) = f(x_j)$,

$$E(X) = x_1 f(x_1) + \cdots + x_n f(x_n) = \sum_{j=1}^{n} x_j f(x_j) = \sum x f(x) \tag{2}$$

where the last summation is taken over all appropriate values of x. As a special case of (2), where the probabilities are all equal, we have

$$E(X) = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{3}$$

which is called the *arithmetic mean*, or simply the *mean*, of $x_1, x_2, \ldots, x_n$.

If X takes on an infinite number of values $x_1, x_2, \ldots$, then $E(X) = \sum_{j=1}^{\infty} x_j f(x_j)$ provided that the infinite series converges absolutely.

For a continuous random variable X having density function $f(x)$, the expectation of X is defined as

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx \tag{4}$$

provided that the integral converges absolutely.

The expectation of X is very often called the *mean* of X and is denoted by μ_X, or simply μ, when the particular random variable is understood.

The mean, or expectation, of X gives a single value that acts as a representative or average of the values of X, and for this reason it is often called a *measure of central tendency.* Other measures are considered on page 83.

EXAMPLE 3.1 Suppose that a game is to be played with a single die assumed fair. In this game a player wins \$20 if a 2 turns up, \$40 if a 4 turns up; loses \$30 if a 6 turns up; while the player neither wins nor loses if any other face turns up. Find the expected sum of money to be won.

Let X be the random variable giving the amount of money won on any toss. The possible amounts won when the die turns up 1, 2, . . . , 6 are $x_1, x_2, \ldots, x_6$, respectively, while the probabilities of these are $f(x_1), f(x_2), \ldots, f(x_6)$. The probability function for X is displayed in Table 3-1. Therefore, the expected value or expectation is

$$E(X) = (0)\left(\frac{1}{6}\right) + (20)\left(\frac{1}{6}\right) + (0)\left(\frac{1}{6}\right) + (40)\left(\frac{1}{6}\right) + (0)\left(\frac{1}{6}\right) + (-30)\left(\frac{1}{6}\right) = 5$$

Table 3-1

x_j	0	+20	0	+40	0	−30
$f(x_j)$	1/6	1/6	1/6	1/6	1/6	1/6

It follows that the player can expect to win \$5. In a fair game, therefore, the player should be expected to pay \$5 in order to play the game.

EXAMPLE 3.2 The density function of a random variable X is given by

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

The expected value of X is then

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = \int_0^2 x\left(\frac{1}{2}x\right)dx = \int_0^2 \frac{x^2}{2}\,dx = \left.\frac{x^3}{6}\right|_0^2 = \frac{4}{3}$$

Functions of Random Variables

Let X be a discrete random variable with probability function $f(x)$. Then $Y = g(X)$ is also a discrete random variable, and the probability function of Y is

$$h(y) = P(Y = y) = \sum_{\{x|g(x)=y\}} P(X = x) = \sum_{\{x|g(x)=y\}} f(x)$$

If X takes on the values $x_1, x_2, \ldots, x_n$, and Y the values $y_1, y_2, \ldots, y_m$ $(m \le n)$, then $y_1h(y_1) + y_2h(y_2) + \cdots + y_mh(y_m) = g(x_1)f(x_1) + g(x_2)f(x_2) + \cdots + g(x_n)f(x_n)$. Therefore,

$$\begin{aligned} E[g(X)] &= g(x_1)f(x_1) + g(x_2)f(x_2) + \cdots + g(x_n)f(x_n) \\ &= \sum_{j=1}^{n} g(x_j)f(x_j) = \sum g(x)f(x) \end{aligned} \tag{5}$$

Similarly, if X is a continuous random variable having probability density $f(x)$, then it can be shown that

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\,dx \tag{6}$$

Note that (5) and (6) do not involve, respectively, the probability function and the probability density function of $Y = g(X)$.

Generalizations are easily made to functions of two or more random variables. For example, if X and Y are two continuous random variables having joint density function $f(x, y)$, then the expectation of $g(X, Y)$ is given by

$$E[g(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)f(x, y)\,dx\,dy \tag{7}$$

EXAMPLE 3.3 If X is the random variable of Example 3.2,

$$E(3X^2 - 2X) = \int_{-\infty}^{\infty} (3x^2 - 2x)f(x)\,dx = \int_0^2 (3x^2 - 2x)\left(\frac{1}{2}x\right)dx = \frac{10}{3}$$

Some Theorems on Expectation

Theorem 3-1 If c is any constant, then

$$E(cX) = cE(X) \tag{8}$$

Theorem 3-2 If X and Y are any random variables, then

$$E(X + Y) = E(X) + E(Y) \tag{9}$$

Theorem 3-3 If X and Y are independent random variables, then

$$E(XY) = E(X)E(Y) \tag{10}$$

Generalizations of these theorems are easily made.

The Variance and Standard Deviation

We have already noted on page 75 that the expectation of a random variable X is often called the *mean* and is denoted by μ. Another quantity of great importance in probability and statistics is called the *variance* and is defined by

$$\text{Var}(X) = E[(X - \mu)^2] \tag{11}$$

The variance is a nonnegative number. The positive square root of the variance is called the *standard deviation* and is given by

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{E[(X - \mu)^2]} \tag{12}$$

Where no confusion can result, the standard deviation is often denoted by σ instead of σ_X, and the variance in such case is σ^2.

If X is a discrete random variable taking the values $x_1, x_2, \ldots, x_n$ and having probability function $f(x)$, then the variance is given by

$$\sigma_X^2 = E[(X - \mu)^2] = \sum_{j=1}^{n}(x_j - \mu)^2 f(x_j) = \sum (x - \mu)^2 f(x) \tag{13}$$

In the special case of (13) where the probabilities are all equal, we have

$$\sigma^2 = [(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_n - \mu)^2]/n \tag{14}$$

which is the variance for a set of n numbers $x_1, \ldots, x_n$.

If X takes on an infinite number of values $x_1, x_2, \ldots$, then $\sigma_X^2 = \sum_{j=1}^{\infty}(x_j - \mu)^2 f(x_j)$, provided that the series converges.

If X is a continuous random variable having density function $f(x)$, then the variance is given by

$$\sigma_X^2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx \tag{15}$$

provided that the integral converges.

The variance (or the standard deviation) is a measure of the *dispersion*, or *scatter*, of the values of the random variable about the mean μ. If the values tend to be concentrated near the mean, the variance is small; while if the values tend to be distributed far from the mean, the variance is large. The situation is indicated graphically in Fig. 3-1 for the case of two continuous distributions having the same mean μ.

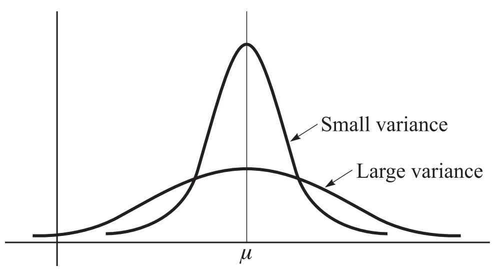


Fig. 3-1

EXAMPLE 3.4 Find the variance and standard deviation of the random variable of Example 3.2. As found in Example 3.2, the mean is $\mu = E(X) = 4/3$. Then the variance is given by

$$\sigma^2 = E\left[\left(X - \frac{4}{3}\right)^2\right] = \int_{-\infty}^{\infty}\left(x - \frac{4}{3}\right)^2 f(x)\,dx = \int_0^2\left(x - \frac{4}{3}\right)^2\left(\frac{1}{2}x\right)dx = \frac{2}{9}$$

and so the standard deviation is $\sigma = \sqrt{\frac{2}{9}} = \frac{\sqrt{2}}{3}$

Note that if X has certain *dimensions* or *units*, such as *centimeters* (cm), then the variance of X has units cm^2 while the standard deviation has the same unit as X, i.e., cm. It is for this reason that the standard deviation is often used.

Some Theorems on Variance

Theorem 3-4
$$\sigma^2 = E[(X - \mu)^2] = E(X^2) - \mu^2 = E(X^2) - [E(X)]^2 \tag{16}$$

where $\mu = E(X)$.

Theorem 3-5 If c is any constant,

$$\mathrm{Var}(cX) = c^2\,\mathrm{Var}(X) \tag{17}$$

Theorem 3-6 The quantity $E[(X - a)^2]$ is a minimum when $a = \mu = E(X)$.

Theorem 3-7 If X and Y are independent random variables,

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \quad \text{or} \quad \sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y \tag{18}$$

$$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \quad \text{or} \quad \sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y \tag{19}$$

Generalizations of Theorem 3-7 to more than two independent variables are easily made. In words, the variance of a sum of independent variables equals the sum of their variances.

Standardized Random Variables

Let X be a random variable with mean μ and standard deviation $\sigma\,(\sigma > 0)$. Then we can define an associated *standardized random variable* given by

$$X^* = \frac{X - \mu}{\sigma} \tag{20}$$

An important property of X^* is that it has a mean of zero and a variance of 1, which accounts for the name *standardized,* i.e.,

$$E(X^*) = 0, \qquad \mathrm{Var}(X^*) = 1 \tag{21}$$

The values of a standardized variable are sometimes called *standard scores,* and X is then said to be expressed in *standard units* (i.e., σ is taken as the unit in measuring $X - \mu$).

Standardized variables are useful for comparing different distributions.

Moments

The *rth moment of a random variable X about the mean* μ, also called the *rth central moment*, is defined as

$$\mu_r = E[(X - \mu)^r] \tag{22}$$

where $r = 0, 1, 2, \ldots$. It follows that $\mu_0 = 1$, $\mu_1 = 0$, and $\mu_2 = \sigma^2$, i.e., the second central moment or second moment about the mean is the variance. We have, assuming absolute convergence,

$$\mu_r = \sum (x - \mu)^r f(x) \qquad \text{(discrete variable)} \tag{23}$$

$$\mu_r = \int_{-\infty}^{\infty} (x - \mu)^r f(x)\, dx \qquad \text{(continuous variable)} \tag{24}$$

The *rth moment of X about the origin*, also called the *rth raw moment*, is defined as

$$\mu_r' = E(X^r) \tag{25}$$

where $r = 0, 1, 2, \ldots$, and in this case there are formulas analogous to (23) and (24) in which $\mu = 0$.

The relationship between these moments is given by

$$\mu_r = \mu_r' - \binom{r}{1}\mu_{r-1}'\mu + \cdots + (-1)^j\binom{r}{j}\mu_{r-j}'\mu^j + \cdots + (-1)^r\mu_0'\mu^r \tag{26}$$

As special cases we have, using $\mu_1' = \mu$ and $\mu_0' = 1$,

$$\begin{aligned}
\mu_2 &= \mu_2' - \mu^2 \\
\mu_3 &= \mu_3' - 3\mu_2'\mu + 2\mu^3 \\
\mu_4 &= \mu_4' - 4\mu_3'\mu + 6\mu_2'\mu^2 - 3\mu^4
\end{aligned} \tag{27}$$

Moment Generating Functions

The *moment generating function* of X is defined by

$$M_X(t) = E(e^{tX}) \tag{28}$$

that is, assuming convergence,

$$M_X(t) = \sum e^{tx} f(x) \qquad \text{(discrete variable)} \tag{29}$$

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx \qquad \text{(continuous variable)} \tag{30}$$

We can show that the Taylor series expansion is [Problem 3.15(a)]

$$M_X(t) = 1 + \mu t + \mu_2' \frac{t^2}{2!} + \cdots + \mu_r' \frac{t^r}{r!} + \cdots \tag{31}$$

Since the coefficients in this expansion enable us to find the moments, the reason for the name *moment generating function* is apparent. From the expansion we can show that [Problem 3.15(b)]

$$\mu_r' = \left.\frac{d^r}{dt^r} M_X(t)\right|_{t=0} \tag{32}$$

i.e., μ_r' is the rth derivative of $M_X(t)$ evaluated at $t = 0$. Where no confusion can result, we often write $M(t)$ instead of $M_X(t)$.

Some Theorems on Moment Generating Functions

Theorem 3-8 If $M_X(t)$ is the moment generating function of the random variable X and a and b $(b \neq 0)$ are constants, then the moment generating function of $(X + a)/b$ is

$$M_{(X+a)/b}(t) = e^{at/b} M_X\left(\frac{t}{b}\right) \tag{33}$$

Theorem 3-9 If X and Y are independent random variables having moment generating functions $M_X(t)$ and $M_Y(t)$, respectively, then

$$M_{X+Y}(t) = M_X(t)\,M_Y(t) \tag{34}$$

Generalizations of Theorem 3-9 to more than two independent random variables are easily made. In words, the moment generating function of a sum of independent random variables is equal to the product of their moment generating functions.

Theorem 3-10 **(Uniqueness Theorem)** Suppose that X and Y are random variables having moment generating functions $M_X(t)$ and $M_Y(t)$, respectively. Then X and Y have the same probability distribution if and only if $M_X(t) = M_Y(t)$ identically.

Characteristic Functions

If we let $t = i\omega$, where i is the imaginary unit, in the moment generating function we obtain an important function called the *characteristic function.* We denote this by

$$\phi_X(\omega) = M_X(i\omega) = E(e^{i\omega X}) \tag{35}$$

It follows that

$$\phi_X(\omega) = \sum e^{i\omega x} f(x) \qquad \text{(discrete variable)} \tag{36}$$

$$\phi_X(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} f(x)\,dx \qquad \text{(continuous variable)} \tag{37}$$

Since $|e^{i\omega x}| = 1$, the series and the integral always converge absolutely.

The corresponding results (31) and (32) become

$$\phi_X(\omega) = 1 + i\mu\omega - \mu_2'\frac{\omega^2}{2!} + \cdots + i^r\mu_r'\frac{\omega^r}{r!} + \cdots \tag{38}$$

where

$$\mu_r' = (-1)^r i^r \frac{d^r}{d\omega^r}\phi_X(\omega)\bigg|_{\omega=0} \tag{39}$$

When no confusion can result, we often write $\phi(\omega)$ instead of $\phi_X(\omega)$.

Theorems for characteristic functions corresponding to Theorems 3-8, 3-9, and 3-10 are as follows.

Theorem 3-11 If $\phi_X(\omega)$ is the characteristic function of the random variable X and a and b $(b \neq 0)$ are constants, then the characteristic function of $(X + a)/b$ is

$$\phi_{(X+a)/b}(\omega) = e^{ai\omega/b}\phi_X\left(\frac{\omega}{b}\right) \tag{40}$$

Theorem 3-12 If X and Y are independent random variables having characteristic functions $\phi_X(\omega)$ and $\phi_Y(\omega)$, respectively, then

$$\phi_{X+Y}(\omega) = \phi_X(\omega)\,\phi_Y(\omega) \tag{41}$$

More generally, the characteristic function of a sum of independent random variables is equal to the product of their characteristic functions.

Theorem 3-13 **(Uniqueness Theorem)** Suppose that X and Y are random variables having characteristic functions $\phi_X(\omega)$ and $\phi_Y(\omega)$, respectively. Then X and Y have the same probability distribution if and only if $\phi_X(\omega) = \phi_Y(\omega)$ identically.

An important reason for introducing the characteristic function is that (37) represents the *Fourier transform* of the density function $f(x)$. From the theory of Fourier transforms, we can easily determine the density function from the characteristic function. In fact,

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega x}\phi_X(\omega)\,d\omega \tag{42}$$

which is often called an *inversion formula*, or *inverse Fourier transform.* In a similar manner we can show in the discrete case that the probability function $f(x)$ can be obtained from (36) by use of *Fourier series*, which is the analog of the Fourier integral for the discrete case. See Problem 3.39.

Another reason for using the characteristic function is that it always exists whereas the moment generating function may not exist.

Variance for Joint Distributions. Covariance

The results given above for one variable can be extended to two or more variables. For example, if X and Y are two continuous random variables having joint density function $f(x, y)$, the means, or expectations, of X and Y are

$$\mu_X = E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x, y)\,dx\,dy, \qquad \mu_Y = E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x, y)\,dx\,dy \tag{43}$$

and the variances are

$$\begin{aligned}\sigma_X^2 &= E[(X - \mu_X)^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - \mu_X)^2 f(x, y)\,dx\,dy \\ \sigma_Y^2 &= E[(Y - \mu_Y)^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y - \mu_Y)^2 f(x, y)\,dx\,dy\end{aligned} \tag{44}$$

Note that the marginal density functions of X and Y are not directly involved in (43) and (44).

Another quantity that arises in the case of two variables X and Y is the *covariance* defined by

$$\sigma_{XY} = \operatorname{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] \tag{45}$$

In terms of the joint density function $f(x, y)$, we have

$$\sigma_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - \mu_X)(y - \mu_Y)f(x, y)\,dx\,dy \tag{46}$$

Similar remarks can be made for two discrete random variables. In such cases (43) and (46) are replaced by

$$\mu_X = \sum_x\sum_y xf(x, y) \qquad \mu_Y = \sum_x\sum_y yf(x, y) \tag{47}$$

$$\sigma_{XY} = \sum_x\sum_y (x - \mu_X)(y - \mu_Y)f(x, y) \tag{48}$$

where the sums are taken over all the discrete values of X and Y.

The following are some important theorems on covariance.

Theorem 3-14

$$\sigma_{XY} = E(XY) - E(X)E(Y) = E(XY) - \mu_X\mu_Y \tag{49}$$

Theorem 3-15 If X and Y are independent random variables, then

$$\sigma_{XY} = \operatorname{Cov}(X, Y) = 0 \tag{50}$$

Theorem 3-16

$$\operatorname{Var}(X \pm Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) \pm 2\operatorname{Cov}(X, Y) \tag{51}$$

or

$$\sigma_{X\pm Y}^2 = \sigma_X^2 + \sigma_Y^2 \pm 2\sigma_{XY} \tag{52}$$

Theorem 3-17

$$|\sigma_{XY}| \le \sigma_X\sigma_Y \tag{53}$$

The converse of Theorem 3-15 is not necessarily true. If X and Y are independent, Theorem 3-16 reduces to Theorem 3-7.

Correlation Coefficient

If X and Y are independent, then $\text{Cov}(X, Y) = \sigma_{XY} = 0$. On the other hand, if X and Y are completely dependent, for example, when $X = Y$, then $\text{Cov}(X, Y) = \sigma_{XY} = \sigma_X \sigma_Y$. From this we are led to a *measure of the dependence* of the variables X and Y given by

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{54}$$

We call ρ the *correlation coefficient*, or *coefficient of correlation.* From Theorem 3-17 we see that $-1 \le \rho \le 1$. In the case where $\rho = 0$ (i.e., the covariance is zero), we call the variables X and Y *uncorrelated.* In such cases, however, the variables may or may not be independent. Further discussion of correlation cases will be given in Chapter 8.

Conditional Expectation, Variance, and Moments

If X and Y have joint density function $f(x, y)$, then as we have seen in Chapter 2, the conditional density function of Y given X is $f(y \mid x) = f(x, y)/f_1(x)$ where $f_1(x)$ is the marginal density function of X. We can define the *conditional expectation*, or *conditional mean*, of Y given X by

$$E(Y \mid X = x) = \int_{-\infty}^{\infty} y f(y \mid x)\, dy \tag{55}$$

where "$X = x$" is to be interpreted as $x < X \le x + dx$ in the continuous case. Theorems 3-1 and 3-2 also hold for conditional expectation.

We note the following properties:

1. $E(Y \mid X = x) = E(Y)$ when X and Y are independent.
2. $E(Y) = \int_{-\infty}^{\infty} E(Y \mid X = x) f_1(x)\, dx$.

It is often convenient to calculate expectations by use of Property 2, rather than directly.

EXAMPLE 3.5 The average travel time to a distant city is c hours by car or b hours by bus. A woman cannot decide whether to drive or take the bus, so she tosses a coin. What is her expected travel time?

Here we are dealing with the joint distribution of the outcome of the toss, X, and the travel time, Y, where $Y = Y_{\text{car}}$ if $X = 0$ and $Y = Y_{\text{bus}}$ if $X = 1$. Presumably, both Y_{car} and Y_{bus} are independent of X, so that by Property 1 above

$$E(Y \mid X = 0) = E(Y_{\text{car}} \mid X = 0) = E(Y_{\text{car}}) = c$$

and

$$E(Y \mid X = 1) = E(Y_{\text{bus}} \mid X = 1) = E(Y_{\text{bus}}) = b$$

Then Property 2 (with the integral replaced by a sum) gives, for a fair coin,

$$E(Y) = E(Y \mid X = 0)P(X = 0) + E(Y \mid X = 1)P(X = 1) = \frac{c + b}{2}$$

In a similar manner we can define the *conditional variance* of Y given X as

$$E[(Y - \mu_2)^2 \mid X = x] = \int_{-\infty}^{\infty} (y - \mu_2)^2 f(y \mid x)\, dy \tag{56}$$

where $\mu_2 = E(Y \mid X = x)$. Also we can define the rth *conditional moment* of Y about any value a given X as

$$E[(Y - a)^r \mid X = x] = \int_{-\infty}^{\infty} (y - a)^r f(y \mid x)\, dy \tag{57}$$

The usual theorems for variance and moments extend to conditional variance and moments.

Chebyshev's Inequality

An important theorem in probability and statistics that reveals a general property of discrete or continuous random variables having finite mean and variance is known under the name of *Chebyshev's inequality.*

Theorem 3-18 **(Chebyshev's Inequality)** Suppose that X is a random variable (discrete or continuous) having mean μ and variance σ^2, which are finite. Then if ϵ is any positive number,

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2} \tag{58}$$

or, with $\epsilon = k\sigma$,

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \tag{59}$$

EXAMPLE 3.6 Letting $k = 2$ in Chebyshev's inequality (59), we see that

$$P(|X - \mu| \geq 2\sigma) \leq 0.25 \qquad \text{or} \qquad P(|X - \mu| < 2\sigma) \geq 0.75$$

In words, the probability of X differing from its mean by more than 2 standard deviations is less than or equal to 0.25; equivalently, the probability that X will lie within 2 standard deviations of its mean is greater than or equal to 0.75. This is quite remarkable in view of the fact that we have not even specified the probability distribution of X.

Law of Large Numbers

The following theorem, called the *law of large numbers*, is an interesting consequence of Chebyshev's inequality.

Theorem 3-19 **(Law of Large Numbers):** Let $X_1, X_2, \ldots, X_n$ be mutually independent random variables (discrete or continuous), each having finite mean μ and variance σ^2. Then if $S_n = X_1 + X_2 + \cdots + X_n (n = 1, 2, \ldots)$,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0 \tag{60}$$

Since S_n/n is the arithmetic mean of $X_1, \ldots, X_n$, this theorem states that the probability of the arithmetic mean S_n/n differing from its expected value μ by more than ϵ approaches zero as $n \to \infty$. A stronger result, which we might expect to be true, is that $\lim_{n\to\infty} S_n/n = \mu$, but this is actually false. However, we can prove that $\lim_{n\to\infty} S_n/n = \mu$ *with probability one.* This result is often called the *strong law of large numbers*, and, by contrast, that of Theorem 3-19 is called the *weak law of large numbers.* When the "law of large numbers" is referred to without qualification, the weak law is implied.

Other Measures of Central Tendency

As we have already seen, the mean, or expectation, of a random variable X provides a measure of central tendency for the values of a distribution. Although the mean is used most, two other measures of central tendency are also employed. These are the *mode* and the *median.*

1. **MODE.** The *mode* of a discrete random variable is that value which occurs most often or, in other words, has the greatest probability of occurring. Sometimes we have two, three, or more values that have relatively large probabilities of occurrence. In such cases, we say that the distribution is *bimodal*, *trimodal*, or *multimodal*, respectively. The mode of a continuous random variable X is the value (or values) of X where the probability density function has a relative maximum.
2. **MEDIAN.** The *median* is that value x for which $P(X < x) \leq \frac{1}{2}$ and $P(X > x) \leq \frac{1}{2}$. In the case of a continuous distribution we have $P(X < x) = \frac{1}{2} = P(X > x)$, and the median separates the density curve into two parts having equal areas of $1/2$ each. In the case of a discrete distribution a unique median may not exist (see Problem 3.34).

Percentiles

It is often convenient to subdivide the area under a density curve by use of ordinates so that the area to the left of the ordinate is some percentage of the total unit area. The values corresponding to such areas are called *percentile values*, or briefly *percentiles.* Thus, for example, the area to the left of the ordinate at x_α in Fig. 3-2 is α. For instance, the area to the left of $x_{0.10}$ would be 0.10, or 10%, and $x_{0.10}$ would be called the *10th percentile* (also called the *first decile*). The median would be the *50th percentile* (or *fifth decile*).

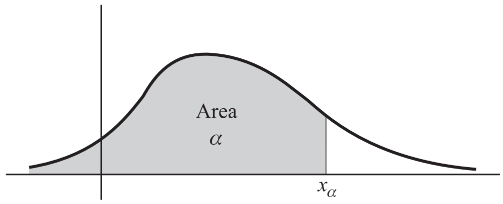


Fig. 3-2

Other Measures of Dispersion

Just as there are various measures of central tendency besides the mean, there are various measures of dispersion or scatter of a random variable besides the variance or standard deviation. Some of the most common are the following.

1. **SEMI-INTERQUARTILE RANGE.** If $x_{0.25}$ and $x_{0.75}$ represent the 25th and 75th percentile values, the difference $x_{0.75} - x_{0.25}$ is called the *interquartile range* and $\frac{1}{2}(x_{0.75} - x_{0.25})$ is the *semi-interquartile range.*
2. **MEAN DEVIATION.** The *mean deviation* (M.D.) of a random variable X is defined as the expectation of $|X - \mu|$, i.e., assuming convergence,

$$\text{M.D.}(X) = E[|X - \mu|] = \sum |x - \mu| f(x) \qquad \text{(discrete variable)} \tag{61}$$

$$\text{M.D.}(X) = E[|X - \mu|] = \int_{-\infty}^{\infty} |x - \mu| f(x)\,dx \qquad \text{(continuous variable)} \tag{62}$$

Skewness and Kurtosis

1. **SKEWNESS.** Often a distribution is not symmetric about any value but instead has one of its tails longer than the other. If the longer tail occurs to the right, as in Fig. 3-3, the distribution is said to be *skewed to the right*, while if the longer tail occurs to the left, as in Fig. 3-4, it is said to be *skewed to the left*. Measures describing this asymmetry are called *coefficients of skewness*, or briefly *skewness*. One such measure is given by

$$\alpha_3 = \frac{E[(X - \mu)^3]}{\sigma^3} = \frac{\mu_3}{\sigma^3} \tag{63}$$

The measure σ_3 will be positive or negative according to whether the distribution is skewed to the right or left, respectively. For a symmetric distribution, $\sigma_3 = 0$.

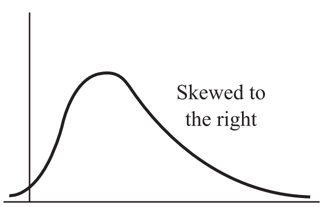


Fig. 3-3

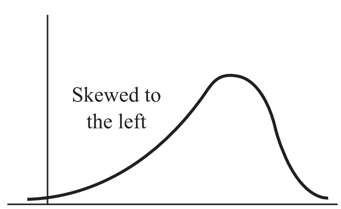


Fig. 3-4

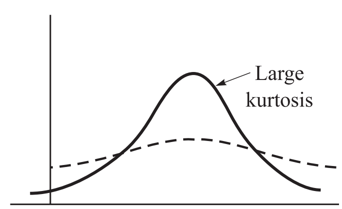


Fig. 3-5

2. **KURTOSIS.** In some cases a distribution may have its values concentrated near the mean so that the distribution has a large peak as indicated by the solid curve of Fig. 3-5. In other cases the distribution may be

relatively flat as in the dashed curve of Fig. 3-5. Measures of the degree of peakedness of a distribution are called *coefficients of kurtosis*, or briefly *kurtosis*. A measure often used is given by

$$\alpha_4 = \frac{E[(X-\mu)^4]}{\sigma^4} = \frac{\mu_4}{\sigma^4} \tag{64}$$

This is usually compared with the normal curve (see Chapter 4), which has a coefficient of kurtosis equal to 3. See also Problem 3.41.

SOLVED PROBLEMS

Expectation of random variables

3.1. In a lottery there are 200 prizes of \$5, 20 prizes of \$25, and 5 prizes of \$100. Assuming that 10,000 tickets are to be issued and sold, what is a fair price to pay for a ticket?

Let X be a random variable denoting the amount of money to be won on a ticket. The various values of X together with their probabilities are shown in Table 3-2. For example, the probability of getting one of the 20 tickets giving a \$25 prize is $20/10{,}000 = 0.002$. The expectation of X in dollars is thus

$$E(X) = (5)(0.02) + (25)(0.002) + (100)(0.0005) + (0)(0.9775) = 0.2$$

or 20 cents. Thus the fair price to pay for a ticket is 20 cents. However, since a lottery is usually designed to raise money, the price per ticket would be higher.

Table 3-2

x (dollars)	5	25	100	0
$P(X = x)$	0.02	0.002	0.0005	0.9775

3.2. Find the expectation of the sum of points in tossing a pair of fair dice.

Let X and Y be the points showing on the two dice. We have

$$E(X) = E(Y) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + \cdots + 6\left(\frac{1}{6}\right) = \frac{7}{2}$$

Then, by Theorem 3-2,

$$E(X + Y) = E(X) + E(Y) = 7$$

3.3. Find the expectation of a discrete random variable X whose probability function is given by

$$f(x) = \left(\frac{1}{2}\right)^x \qquad (x = 1, 2, 3, \ldots)$$

We have

$$E(X) = \sum_{x=1}^{\infty} x\left(\frac{1}{2}\right)^x = \frac{1}{2} + 2\left(\frac{1}{4}\right) + 3\left(\frac{1}{8}\right) + \cdots$$

To find this sum, let

$$S = \frac{1}{2} + 2\left(\frac{1}{4}\right) + 3\left(\frac{1}{8}\right) + 4\left(\frac{1}{16}\right) + \cdots$$

Then

$$\frac{1}{2}S = \frac{1}{4} + 2\left(\frac{1}{8}\right) + 3\left(\frac{1}{16}\right) + \cdots$$

Subtracting,

$$\frac{1}{2}S = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = 1$$

Therefore, $S = 2$.

3.4. A continuous random variable X has probability density given by

$$f(x) = \begin{cases} 2e^{-2x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Find (a) $E(X)$, (b) $E(X^2)$.

(a)
$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = \int_0^{\infty} x(2e^{-2x})\,dx = 2\int_0^{\infty} xe^{-2x}\,dx$$
$$= 2\left[(x)\left(\frac{e^{-2x}}{-2}\right) - (1)\left(\frac{e^{-2x}}{4}\right)\right]\Bigg|_0^{\infty} = \frac{1}{2}$$

(b)
$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = 2\int_0^{\infty} x^2 e^{-2x}\,dx$$
$$= 2\left[(x^2)\left(\frac{e^{-2x}}{-2}\right) - (2x)\left(\frac{e^{-2x}}{4}\right) + (2)\left(\frac{e^{-2x}}{-8}\right)\right]\Bigg|_0^{\infty} = \frac{1}{2}$$

3.5. The joint density function of two random variables X and Y is given by

$$f(x, y) = \begin{cases} xy/96 & 0 < x < 4,\ 1 < y < 5 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) $E(X)$, (b) $E(Y)$, (c) $E(XY)$, (d) $E(2X + 3Y)$.

(a)
$$E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x, y)\,dx\,dy = \int_{x=0}^{4}\int_{y=1}^{5} x\left(\frac{xy}{96}\right)dx\,dy = \frac{8}{3}$$

(b)
$$E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x, y)\,dx\,dy = \int_{x=0}^{4}\int_{y=1}^{5} y\left(\frac{xy}{96}\right)dx\,dy = \frac{31}{9}$$

(c)
$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (xy)f(x, y)\,dx\,dy = \int_{x=0}^{4}\int_{y=1}^{5} (xy)\left(\frac{xy}{96}\right)dx\,dy = \frac{248}{27}$$

(d)
$$E(2X + 3Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (2x + 3y)f(x, y)\,dx\,dy = \int_{x=0}^{4}\int_{y=1}^{5} (2x + 3y)\left(\frac{xy}{96}\right)dx\,dy = \frac{47}{3}$$

Another method

(c) Since X and Y are independent, we have, using parts (a) and (b),

$$E(XY) = E(X)E(Y) = \left(\frac{8}{3}\right)\left(\frac{31}{9}\right) = \frac{248}{27}$$

(d) By Theorems 3-1 and 3-2, pages 76–77, together with (a) and (b),

$$E(2X + 3Y) = 2E(X) + 3E(Y) = 2\left(\frac{8}{3}\right) + 3\left(\frac{31}{9}\right) = \frac{47}{3}$$

3.6. Prove Theorem 3-2, page 77.

Let $f(x, y)$ be the joint probability function of X and Y, assumed discrete. Then

$$\begin{aligned} E(X + Y) &= \sum_x \sum_y (x + y)f(x, y) \\ &= \sum_x \sum_y xf(x, y) + \sum_x \sum_y yf(x, y) \\ &= E(X) + E(Y) \end{aligned}$$

If either variable is continuous, the proof goes through as before, with the appropriate summations replaced by integrations. Note that the theorem is true whether or not X and Y are independent.

3.7. Prove Theorem 3-3, page 77.

Let $f(x, y)$ be the joint probability function of X and Y, assumed discrete. If the variables X and Y are independent, we have $f(x, y) = f_1(x) f_2(y)$. Therefore,

$$\begin{aligned} E(XY) &= \sum_x \sum_y xyf(x, y) = \sum_x \sum_y xyf_1(x)f_2(y) \\ &= \sum_x \left[xf_1(x) \sum_y yf_2(y) \right] \\ &= \sum_x [(xf_1(x)E(y)] \\ &= E(X)E(Y) \end{aligned}$$

If either variable is continuous, the proof goes through as before, with the appropriate summations replaced by integrations. Note that the validity of this theorem hinges on whether $f(x, y)$ can be expressed as a function of x multiplied by a function of y, for all x and y, i.e., on whether X and Y are independent. For dependent variables it is not true in general.

Variance and standard deviation

3.8. Find (a) the variance, (b) the standard deviation of the sum obtained in tossing a pair of fair dice.

(a) Referring to Problem 3.2, we have $E(X) = E(Y) = 1/2$. Moreover,

$$E(X^2) = E(Y^2) = 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + \cdots + 6^2\left(\frac{1}{6}\right) = \frac{91}{6}$$

Then, by Theorem 3-4,

$$\mathrm{Var}(X) = \mathrm{Var}(Y) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

and, since X and Y are independent, Theorem 3-7 gives

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = \frac{35}{6}$$

(b)
$$\sigma_{X+Y} = \sqrt{\mathrm{Var}(X + Y)} = \sqrt{\frac{35}{6}}$$

3.9. Find (a) the variance, (b) the standard deviation for the random variable of Problem 3.4.

(a) As in Problem 3.4, the mean of X is $\mu = E(X) = \frac{1}{2}$. Then the variance is

$$\mathrm{Var}(X) = E[(X - \mu)^2] = E\left[\left(X - \frac{1}{2}\right)^2\right] = \int_{-\infty}^{\infty} \left(x - \frac{1}{2}\right)^2 f(x)\,dx$$

$$= \int_0^{\infty} \left(x - \frac{1}{2}\right)^2 (2e^{-2x})\,dx = \frac{1}{4}$$

Another method
By Theorem 3-4,

$$\mathrm{Var}(X) = E[(X - \mu)^2] = E(X^2) - [E(X)]^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

(b)
$$\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{\frac{1}{4}} = \frac{1}{2}$$

3.10. Prove Theorem 3-4, page 78.

We have

$$
\begin{aligned}
E[(X-\mu)^2] &= E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2 \\
&= E(X^2) - [E(X)]^2
\end{aligned}
$$

3.11. Prove Theorem 3-6, page 78.

$$
\begin{aligned}
E[(X-a)^2] &= E[\{(X-\mu) + (\mu - a)\}^2] \\
&= E[(X-\mu)^2 + 2(X-\mu)(\mu-a) + (\mu-a)^2] \\
&= E[(X-\mu)^2] + 2(\mu-a)E(X-\mu) + (\mu-a)^2 \\
&= E[(X-\mu)^2] + (\mu-a)^2
\end{aligned}
$$

since $E(X-\mu) = E(X) - \mu = 0$. From this we see that the minimum value of $E[(X-a)^2]$ occurs when $(\mu - a)^2 = 0$, i.e., when $a = \mu$.

3.12. If $X^* = (X-\mu)/\sigma$ is a standardized random variable, prove that (a) $E(X^*) = 0$, (b) $\mathrm{Var}(X^*) = 1$.

(a)
$$E(X^*) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}[E(X-\mu)] = \frac{1}{\sigma}[E(X) - \mu] = 0$$

since $E(X) = \mu$.

(b)
$$\mathrm{Var}(X^*) = \mathrm{Var}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}E[(X-\mu)^2] = 1$$

using Theorem 3-5, page 78, and the fact that $E[(X-\mu)^2] = \sigma^2$.

3.13. Prove Theorem 3-7, page 78.

$$
\begin{aligned}
\mathrm{Var}(X+Y) &= E[\{(X+Y) - (\mu_X + \mu_Y)\}^2] \\
&= E[\{(X-\mu_X) + (Y-\mu_Y)\}^2] \\
&= E[(X-\mu_X)^2 + 2(X-\mu_X)(Y-\mu_Y) + (Y-\mu_Y)^2] \\
&= E[(X-\mu_X)^2] + 2E[(X-\mu_X)(Y-\mu_Y)] + E[(Y-\mu_Y)^2] \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y)
\end{aligned}
$$

using the fact that

$$E[(X-\mu_X)(Y-\mu_Y)] = E(X-\mu_X)E(Y-\mu_Y) = 0$$

since X and Y, and therefore $X - \mu_X$ and $Y - \mu_Y$, are independent. The proof of (*19*), page 78, follows on replacing Y by $-Y$ and using Theorem 3-5.

Moments and moment generating functions

3.14. Prove the result (26), page 79.

$$
\begin{aligned}
\mu_r &= E[(X-\mu)^r] \\
&= E\left[X^r - \binom{r}{1}X^{r-1}\mu + \cdots + (-1)^j\binom{r}{j}X^{r-j}\mu^j \right. \\
&\qquad \left. + \cdots + (-1)^{r-1}\binom{r}{r-1}X\mu^{r-1} + (-1)^r\mu^r\right]
\end{aligned}
$$

$$= E(X^r) - \binom{r}{1}E(X^{r-1})\mu + \cdots + (-1)^j\binom{r}{j}E(X^{r-j})\mu^j$$

$$+ \cdots + (-1)^{r-1}\binom{r}{r-1}E(X)\mu^{r-1} + (-1)^r\mu^r$$

$$= \mu'_r - \binom{r}{1}\mu'_{r-1}\mu + \cdots + (-1)^j\binom{r}{j}\mu'_{r-j}\mu^j$$

$$+ \cdots + (-1)^{r-1}r\mu^r + (-1)^{-r}\mu^r$$

where the last two terms can be combined to give $(-1)^{r-1}(r-1)\mu^r$.

3.15. Prove (a) result (31), (b) result (32), page 79.

(a) Using the power series expansion for e^u (3., Appendix A), we have

$$M_X(t) = E(e^{tX}) = E\left(1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \cdots\right)$$

$$= 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \frac{t^3}{3!}E(X^3) + \cdots$$

$$= 1 + \mu t + \mu'_2\frac{t^2}{2!} + \mu'_3\frac{t^3}{3!} + \cdots$$

(b) This follows immediately from the fact known from calculus that if the Taylor series of $f(t)$ about $t = a$ is

$$f(t) = \sum_{n=0}^{\infty} c_n(t-a)^n$$

then
$$c_n = \frac{1}{n!}\frac{d^n}{dt^n}f(t)\bigg|_{t=a}$$

3.16. Prove Theorem 3-9, page 80.

Since X and Y are independent, any function of X and any function of Y are independent. Hence,

$$M_{X+Y}(t) = E[e^{t(X+Y)}] = E(e^{tX}e^{tY}) = E(e^{tX})E(e^{tY}) = M_X(t)M_Y(t)$$

3.17. The random variable X can assume the values 1 and -1 with probability $\frac{1}{2}$ each. Find (a) the moment generating function, (b) the first four moments about the origin.

(a)
$$E(e^{tX}) = e^{t(1)}\left(\frac{1}{2}\right) + e^{t(-1)}\left(\frac{1}{2}\right) = \frac{1}{2}(e^t + e^{-t})$$

(b) We have
$$e^t = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \cdots$$

$$e^{-t} = 1 - t + \frac{t^2}{2!} - \frac{t^3}{3!} + \frac{t^4}{4!} - \cdots$$

Then (1)
$$\frac{1}{2}(e^t + e^{-t}) = 1 + \frac{t^2}{2!} + \frac{t^4}{4!} + \cdots$$

But (2)
$$M_X(t) = 1 + \mu t + \mu'_2\frac{t^2}{2!} + \mu'_3\frac{t^3}{3!} + \mu'_4\frac{t^4}{4!} + \cdots$$

Then, comparing (1) and (2), we have

$$\mu = 0, \qquad \mu'_2 = 1, \qquad \mu'_3 = 0, \qquad \mu'_4 = 1, \ldots$$

The odd moments are all zero, and the even moments are all one.

3.18. A random variable X has density function given by

$$f(x) = \begin{cases} 2e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Find (a) the moment generating function, (b) the first four moments about the origin.

(a)
$$M(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

$$= \int_0^{\infty} e^{tx}(2e^{-2x})\,dx = 2\int_0^{\infty} e^{(t-2)x}\,dx$$

$$= \left.\frac{2e^{(t-2)x}}{t-2}\right|_0^{\infty} = \frac{2}{2-t}, \quad \text{assuming } t < 2$$

(b) If $|t| < 2$ we have

$$\frac{2}{2-t} = \frac{1}{1-t/2} = 1 + \frac{t}{2} + \frac{t^2}{4} + \frac{t^3}{8} + \frac{t^4}{16} + \cdots$$

But
$$M(t) = 1 + \mu t + \mu_2' \frac{t^2}{2!} + \mu_3' \frac{t^3}{3!} + \mu_4' \frac{t^4}{4!} + \cdots$$

Therefore, on comparing terms, $\mu = \frac{1}{2}, \mu_2' = \frac{1}{2}, \mu_3' = \frac{3}{4}, \mu_4' = \frac{3}{2}$.

3.19. Find the first four moments (a) about the origin, (b) about the mean, for a random variable X having density function

$$f(x) = \begin{cases} 4x(9 - x^2)/81 & 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

(a)
$$\mu_1' = E(X) = \frac{4}{81}\int_0^3 x^2(9 - x^2)\,dx = \frac{8}{5} = \mu$$

$$\mu_2' = E(X^2) = \frac{4}{81}\int_0^3 x^3(9 - x^2)\,dx = 3$$

$$\mu_3' = E(X^3) = \frac{4}{81}\int_0^3 x^4(9 - x^2)\,dx = \frac{216}{35}$$

$$\mu_4' = E(X^4) = \frac{4}{81}\int_0^3 x^5(9 - x^2)\,dx = \frac{27}{2}$$

(b) Using the result (27), page 79, we have

$$\mu_1 = 0$$

$$\mu_2 = 3 - \left(\frac{8}{5}\right)^2 = \frac{11}{25} = \sigma^2$$

$$\mu_3 = \frac{216}{35} - 3(3)\left(\frac{8}{5}\right) + 2\left(\frac{8}{5}\right)^3 = -\frac{32}{875}$$

$$\mu_4 = \frac{27}{2} - 4\left(\frac{216}{35}\right)\left(\frac{8}{5}\right) + 6(3)\left(\frac{8}{5}\right)^2 - 3\left(\frac{8}{5}\right)^4 = \frac{3693}{8750}$$

Characteristic functions

3.20. Find the characteristic function of the random variable X of Problem 3.17.

The characteristic function is given by

$$E(e^{i\omega X}) = e^{i\omega(1)}\left(\frac{1}{2}\right) + e^{i\omega(-1)}\left(\frac{1}{2}\right) = \frac{1}{2}(e^{i\omega} + e^{-i\omega}) = \cos\omega$$

using Euler's formulas,

$$e^{i\theta} = \cos\theta + i\sin\theta \qquad e^{-i\theta} = \cos\theta - i\sin\theta$$

with $\theta = \omega$. The result can also be obtained from Problem 3.17(a) on putting $t = i\omega$.

3.21. Find the characteristic function of the random variable X having density function given by

$$f(x) = \begin{cases} 1/2a & |x| < a \\ 0 & \text{otherwise} \end{cases}$$

The characteristic function is given by

$$\begin{aligned} E(e^{i\omega X}) &= \int_{-\infty}^{\infty} e^{i\omega x} f(x)\,dx = \frac{1}{2a}\int_{-a}^{a} e^{i\omega x}\,dx \\ &= \frac{1}{2a}\frac{e^{i\omega x}}{i\omega}\bigg|_{-a}^{a} = \frac{e^{ia\omega} - e^{-ia\omega}}{2ia\omega} = \frac{\sin a\omega}{a\omega} \end{aligned}$$

using Euler's formulas (see Problem 3.20) with $\theta = a\omega$.

3.22. Find the characteristic function of the random variable X having density function $f(x) = ce^{-a|x|}$, $-\infty < x < \infty$, where $a > 0$, and c is a suitable constant.

Since $f(x)$ is a density function, we must have

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

so that

$$\begin{aligned} c\int_{-\infty}^{\infty} e^{-a|x|}\,dx &= c\left[\int_{-\infty}^{0} e^{-a(-x)}\,dx + \int_{0}^{\infty} e^{-a(x)}\,dx\right] \\ &= c\frac{e^{ax}}{a}\bigg|_{-\infty}^{0} + c\frac{e^{-ax}}{-a}\bigg|_{0}^{\infty} = \frac{2c}{a} = 1 \end{aligned}$$

Then $c = a/2$. The characteristic function is therefore given by

$$\begin{aligned} E(e^{i\omega X}) &= \int_{-\infty}^{\infty} e^{i\omega x} f(x)\,dx \\ &= \frac{a}{2}\left[\int_{-\infty}^{0} e^{i\omega x}e^{-a(-x)}\,dx + \int_{0}^{\infty} e^{i\omega x}e^{-a(x)}\,dx\right] \\ &= \frac{a}{2}\left[\int_{-\infty}^{0} e^{(a+i\omega)x}\,dx + \int_{0}^{\infty} e^{-(a-i\omega)x}\,dx\right] \\ &= \frac{a}{2}\frac{e^{(a+i\omega)x}}{a + i\omega}\bigg|_{-\infty}^{0} + a\frac{e^{-(a-i\omega)x}}{-(a - i\omega)}\bigg|_{0}^{\infty} \\ &= \frac{a}{2(a + i\omega)} + \frac{a}{2(a - i\omega)} = \frac{a^2}{a^2 + \omega^2} \end{aligned}$$

Covariance and correlation coefficient

3.23. Prove Theorem 3-14, page 81.

By definition the covariance of X and Y is

$$\begin{aligned} \sigma_{XY} &= \mathrm{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y] \\ &= E(XY) - \mu_X E(Y) - \mu_Y E(X) + E(\mu_X\mu_Y) \\ &= E(XY) - \mu_X\mu_Y - \mu_Y\mu_X + \mu_X\mu_Y \\ &= E(XY) - \mu_X\mu_Y \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

3.24. Prove Theorem 3-15, page 81.

If X and Y are independent, then $E(XY) = E(X)E(Y)$. Therefore, by Problem 3.23,

$$\sigma_{XY} = \text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$$

3.25. Find (a) $E(X)$, (b) $E(Y)$, (c) $E(XY)$, (d) $E(X^2)$, (e) $E(Y^2)$, (f) Var (X), (g) Var (Y), (h) Cov (X, Y), (i) ρ, if the random variables X and Y are defined as in Problem 2.8, pages 47–48.

(a)
$$E(X) = \sum_x \sum_y x f(x, y) = \sum_x x\left[\sum_y f(x, y)\right]$$
$$= (0)(6c) + (1)(14c) + (2)(22c) = 58c = \frac{58}{42} = \frac{29}{21}$$

(b)
$$E(Y) = \sum_x \sum_y y f(x, y) = \sum_y y\left[\sum_x f(x, y)\right]$$
$$= (0)(6c) + (1)(9c) + (2)(12c) + (3)(15c) = 78c = \frac{78}{42} = \frac{13}{7}$$

(c)
$$E(XY) = \sum_x \sum_y xy f(x, y)$$
$$= (0)(0)(0) + (0)(1)(c) + (0)(2)(2c) + (0)(3)(3c)$$
$$+ (1)(0)(2c) + (1)(1)(3c) + (1)(2)(4c) + (1)(3)(5c)$$
$$+ (2)(0)(4c) + (2)(1)(5c) + (2)(2)(6c) + (2)(3)(7c)$$
$$= 102c = \frac{102}{42} = \frac{17}{7}$$

(d)
$$E(X^2) = \sum_x \sum_y x^2 f(x, y) = \sum_x x^2\left[\sum_y f(x, y)\right]$$
$$= (0)^2(6c) + (1)^2(14c) + (2)^2(22c) = 102c = \frac{102}{42} = \frac{17}{7}$$

(e)
$$E(Y^2) = \sum_x \sum_y y^2 f(x, y) = \sum_y y^2\left[\sum_x f(x, y)\right]$$
$$= (0)^2(6c) + (1)^2(9c) + (2)^2(12c) + (3)^2(15c) = 192c = \frac{192}{42} = \frac{32}{7}$$

(f)
$$\sigma_X^2 = \text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{17}{7} - \left(\frac{29}{21}\right)^2 = \frac{230}{441}$$

(g)
$$\sigma_Y^2 = \text{Var}(Y) = E(Y^2) - [E(Y)]^2 = \frac{32}{7} - \left(\frac{13}{7}\right)^2 = \frac{55}{49}$$

(h)
$$\sigma_{XY} = \text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{17}{7} - \left(\frac{29}{21}\right)\left(\frac{13}{7}\right) = -\frac{20}{147}$$

(i)
$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{-20/147}{\sqrt{230/441}\sqrt{55/49}} = \frac{-20}{\sqrt{230}\sqrt{55}} = -0.2103 \text{ approx.}$$

3.26. Work Problem 3.25 if the random variables X and Y are defined as in Problem 2.33, pages 61–63.

Using $c = 1/210$, we have:

(a)
$$E(X) = \frac{1}{210}\int_{x=2}^{6}\int_{y=0}^{5} (x)(2x + y)\, dx\, dy = \frac{268}{63}$$

(b)
$$E(Y) = \frac{1}{210}\int_{x=2}^{6}\int_{y=0}^{5} (y)(2x + y)\, dx\, dy = \frac{170}{63}$$

(c)
$$E(XY) = \frac{1}{210}\int_{x=2}^{6}\int_{y=0}^{5} (xy)(2x + y)\, dx\, dy = \frac{80}{7}$$

(d) $$E(X^2) = \frac{1}{210}\int_{x=2}^{6}\int_{y=0}^{5}(x^2)(2x + y)\,dx\,dy = \frac{1220}{63}$$

(e) $$E(Y^2) = \frac{1}{210}\int_{x=2}^{6}\int_{y=0}^{5}(y^2)(2x + y)\,dx\,dy = \frac{1175}{126}$$

(f) $$\sigma_X^2 = \operatorname{Var}(X) = E(X^2) - [E(X)]^2 = \frac{1220}{63} - \left(\frac{268}{63}\right)^2 = \frac{5036}{3969}$$

(g) $$\sigma_Y^2 = \operatorname{Var}(Y) = E(Y^2) - [E(Y)]^2 = \frac{1175}{126} - \left(\frac{170}{63}\right)^2 = \frac{16{,}225}{7938}$$

(h) $$\sigma_{XY} = \operatorname{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{80}{7} - \left(\frac{268}{63}\right)\left(\frac{170}{63}\right) = -\frac{200}{3969}$$

(i) $$\rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{-200/3969}{\sqrt{5036/3969}\sqrt{16{,}225/7938}} = \frac{-200}{\sqrt{2518}\sqrt{16{,}225}} = -0.03129 \text{ approx.}$$

Conditional expectation, variance, and moments

3.27. Find the conditional expectation of Y given $X = 2$ in Problem 2.8, pages 47–48.

As in Problem 2.27, page 58, the conditional probability function of Y given $X = 2$ is

$$f(y\,|\,2) = \frac{4 + y}{22}$$

Then the conditional expectation of Y given $X = 2$ is

$$E(Y\,|\,X = 2) = \sum_y y\left(\frac{4 + y}{22}\right)$$

where the sum is taken over all y corresponding to $X = 2$. This is given by

$$E(Y\,|\,X = 2) = (0)\left(\frac{4}{22}\right) + 1\left(\frac{5}{22}\right) + 2\left(\frac{6}{22}\right) + 3\left(\frac{7}{22}\right) = \frac{19}{11}$$

3.28. Find the conditional expectation of (a) Y given X, (b) X given Y in Problem 2.29, pages 58–59.

(a) $$E(Y\,|\,X = x)\int_{-\infty}^{\infty} yf_2(y\,|\,x)\,dy = \int_0^x y\left(\frac{2y}{x^2}\right)dy = \frac{2x}{3}$$

(b) $$E(X\,|\,Y = y) = \int_{-\infty}^{\infty} xf_1(x\,|\,y)\,dx = \int_y^1 x\left(\frac{2x}{1 - y^2}\right)dx$$

$$= \frac{2(1 - y^3)}{3(1 - y^2)} = \frac{2(1 + y + y^2)}{3(1 + y)}$$

3.29. Find the conditional variance of Y given X for Problem 2.29, pages 58–59.

The required variance (second moment about the mean) is given by

$$E[(Y - \mu_2)^2\,|\,X = x] = \int_{-\infty}^{\infty}(y - \mu_2)^2 f_2(y\,|\,x)\,dy = \int_0^x\left(y - \frac{2x}{3}\right)^2\left(\frac{2y}{x^2}\right)dy = \frac{x^2}{18}$$

where we have used the fact that $\mu_2 = E(Y\,|\,X = x) = 2x/3$ from Problem 3.28(a).

Chebyshev's inequality

3.30. Prove Chebyshev's inequality.

We shall present the proof for continuous random variables. A proof for discrete variables is similar if integrals are replaced by sums. If $f(x)$ is the density function of X, then

$$\sigma^2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty}(x - \mu)^2 f(x)\,dx$$

Since the integrand is nonnegative, the value of the integral can only decrease when the range of integration is diminished. Therefore,

$$\sigma^2 \geq \int_{|x-\mu|\geq\epsilon} (x-\mu)^2 f(x)\,dx \geq \int_{|x-\mu|\geq\epsilon} \epsilon^2 f(x)\,dx = \epsilon^2 \int_{|x-\mu|\geq\epsilon} f(x)\,dx$$

But the last integral is equal to $P(|X-\mu| \geq \epsilon)$. Hence,

$$P(|X-\mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

3.31. For the random variable of Problem 3.18, (a) find $P(|X-\mu| > 1)$. (b) Use Chebyshev's inequality to obtain an upper bound on $P(|X-\mu| > 1)$ and compare with the result in (a).

(a) From Problem 3.18, $\mu = 1/2$. Then

$$P(|X-\mu| < 1) = P\left(\left|X - \frac{1}{2}\right| < 1\right) = P\left(-\frac{1}{2} < X < \frac{3}{2}\right)$$

$$= \int_0^{3/2} 2e^{-2x}\,dx = 1 - e^{-3}$$

Therefore $$P\left(\left|X - \frac{1}{2}\right| \geq 1\right) = 1 - (1 - e^{-3}) = e^{-3} = 0.04979$$

(b) From Problem 3.18, $\sigma^2 = \mu_2' - \mu^2 = 1/4$. Chebyshev's inequality with $\epsilon = 1$ then gives

$$P(|X-\mu| \geq 1) \leq \sigma^2 = 0.25$$

Comparing with (a), we see that the bound furnished by Chebyshev's inequality is here quite crude. In practice, Chebyshev's inequality is used to provide estimates when it is inconvenient or impossible to obtain exact values.

Law of large numbers

3.32. Prove the law of large numbers stated in Theorem 3-19, page 83.

We have $$E(X_1) = E(X_2) = \cdots = E(X_n) = \mu$$

$$\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \cdots = \mathrm{Var}(X_n) = \sigma^2$$

Then $$E\left(\frac{S_n}{n}\right) = E\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n}[E(X_1) + \cdots + E(X_n)] = \frac{1}{n}(n\mu) = \mu$$

$$\mathrm{Var}(S_n) = \mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = n\sigma^2$$

so that $$\mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\mathrm{Var}(S_n) = \frac{\sigma^2}{n}$$

where we have used Theorem 3-5 and an extension of Theorem 3-7.

Therefore, by Chebyshev's inequality with $X = S_n/n$, we have

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}$$

Taking the limit as $n \to \infty$, this becomes, as required,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0$$

Other measures of central tendency

3.33. The density function of a continuous random variable X is

$$f(x) = \begin{cases} 4x(9-x^2)/81 & 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the mode. (b) Find the median. (c) Compare mode, median, and mean.

(a) The mode is obtained by finding where the density $f(x)$ has a relative maximum. The relative maxima of $f(x)$ occur where the derivative is zero, i.e.,

$$\frac{d}{dx}\left[\frac{4x(9-x^2)}{81}\right] = \frac{36-12x^2}{81} = 0$$

Then $x = \sqrt{3} = 1.73$ approx., which is the required mode. Note that this does give the maximum since the second derivative, $-24x/81$, is negative for $x = \sqrt{3}$.

(b) The median is that value a for which $P(X \le a) = 1/2$. Now, for $0 < a < 3$,

$$P(X \le a) = \frac{4}{81}\int_0^a x(9-x^2)\,dx = \frac{4}{81}\left(\frac{9a^2}{2} - \frac{a^4}{4}\right)$$

Setting this equal to $1/2$, we find that

$$2a^4 - 36a^2 + 81 = 0$$

from which

$$a^2 = \frac{36 \pm \sqrt{(36)^2 - 4(2)(81)}}{2(2)} = \frac{36 \pm \sqrt{648}}{4} = 9 \pm \frac{9}{2}\sqrt{2}$$

Therefore, the required median, which must lie between 0 and 3, is given by

$$a^2 = 9 - \frac{9}{2}\sqrt{2}$$

from which $a = 1.62$ approx.

(c)
$$E(X) = \frac{4}{81}\int_0^3 x^2(9-x^2)\,dx = \frac{4}{81}\left(3x^3 - \frac{x^5}{5}\right)\Bigg|_0^3 = 1.60$$

which is practically equal to the median. The mode, median, and mean are shown in Fig. 3-6.

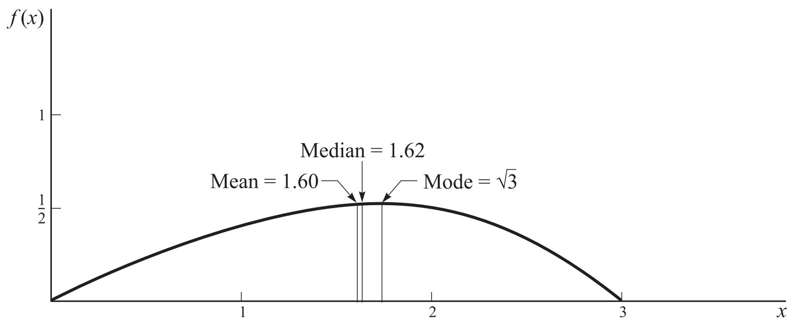


Fig. 3-6

3.34. A discrete random variable has probability function $f(x) = 1/2^x$ where $x = 1, 2, \ldots$. Find (a) the mode, (b) the median, and (c) compare them with the mean.

(a) The mode is the value x having largest associated probability. In this case it is $x = 1$, for which the probability is $1/2$.

(b) If x is any value between 1 and 2, $P(X < x) = \frac{1}{2}$ and $P(X > x) = \frac{1}{2}$. Therefore, *any number* between 1 and 2 could represent the median. For convenience, we choose the midpoint of the interval, i.e., $3/2$.

(c) As found in Problem 3.3, $\mu = 2$. Therefore, the ordering of the three measures is just the reverse of that in Problem 3.33.

Percentiles

3.35. Determine the (a) 10th, (b) 25th, (c) 75th percentile values for the distribution of Problem 3.33.

From Problem 3.33(b) we have

$$P(X \leq a) = \frac{4}{81}\left(\frac{9a^2}{2} - \frac{a^4}{4}\right) = \frac{18a^2 - a^4}{81}$$

(a) The 10th percentile is the value of a for which $P(X \leq a) = 0.10$, i.e., the solution of $(18a^2 - a^4)/81 = 0.10$. Using the method of Problem 3.33, we find $a = 0.68$ approx.

(b) The 25th percentile is the value of a such that $(18a^2 - a^4)/81 = 0.25$, and we find $a = 1.098$ approx.

(c) The 75th percentile is the value of a such that $(18a^2 - a^4)/81 = 0.75$, and we find $a = 2.121$ approx.

Other measures of dispersion

3.36. Determine, (a) the semi-interquartile range, (b) the mean deviation for the distribution of Problem 3.33.

(a) By Problem 3.35 the 25th and 75th percentile values are 1.098 and 2.121, respectively. Therefore,

$$\text{Semi-interquartile range} = \frac{2.121 - 1.098}{2} = 0.51 \text{ approx.}$$

(b) From Problem 3.33 the mean is $\mu = 1.60 = 8/5$. Then

$$\begin{aligned}
\text{Mean deviation} = \text{M.D.} = E(|X - \mu|) &= \int_{-\infty}^{\infty} |x - \mu| f(x)\, dx \\
&= \int_0^3 \left|x - \frac{8}{5}\right| \left[\frac{4x}{81}(9 - x^2)\right] dx \\
&= \int_0^{8/5} \left(\frac{8}{5} - x\right)\left[\frac{4x}{81}(9 - x^2)\right] dx + \int_{8/5}^{3} \left(x - \frac{8}{5}\right)\left[\frac{4x}{81}(9 - x^2)\right] dx \\
&= 0.555 \text{ approx.}
\end{aligned}$$

Skewness and kurtosis

3.37. Find the coefficient of (a) skewness, (b) kurtosis for the distribution of Problem 3.19.

From Problem 3.19(b) we have

$$\sigma^2 = \frac{11}{25} \qquad \mu_3 = -\frac{32}{875} \qquad \mu_4 = \frac{3693}{8750}$$

(a) Coefficient of skewness $= \alpha_3 = \dfrac{\mu_3}{\sigma^3} = -0.1253$

(b) Coefficient of kurtosis $= \alpha_4 = \dfrac{\mu_4}{\sigma^4} = 2.172$

It follows that there is a moderate skewness to the left, as is indicated in Fig. 3-6. Also the distribution is somewhat less peaked than the normal distribution, which has a kurtosis of 3.

Miscellaneous problems

3.38. If $M(t)$ is the moment generating function for a random variable X, prove that the mean is $\mu = M'(0)$ and the variance is $\sigma^2 = M''(0) - [M'(0)]^2$.

From (32), page 79, we have on letting $r = 1$ and $r = 2$,

$$\mu_1' = M'(0) \qquad \mu_2' = M''(0)$$

Then from (27)

$$\mu = M'(0) \qquad \mu_2 = \sigma^2 = M''(0) - [M'(0)]^2$$

3.39. Let X be a random variable that takes on the values $x_k = k$ with probabilities p_k where $k = \pm 1, \ldots, \pm n$. (a) Find the characteristic function $\phi(\omega)$ of X, (b) obtain p_k in terms of $\phi(\omega)$.

(a) The characteristic function is

$$\phi(\omega) = E(e^{i\omega X}) = \sum_{k=-n}^{n} e^{i\omega x_k} p_k = \sum_{k=-n}^{n} p_k e^{ik\omega}$$

(b) Multiply both sides of the expression in (a) by $e^{-ij\omega}$ and integrate with respect to ω from 0 to 2π. Then

$$\int_{\omega=0}^{2\pi} e^{-ij\omega}\phi(\omega)\,d\omega = \sum_{k=-n}^{n} p_k \int_{\omega=0}^{2\pi} e^{i(k-j)\omega}\,d\omega = 2\pi p_j$$

since

$$\int_{\omega=0}^{2\pi} e^{i(k-j)\omega}\,d\omega = \begin{cases} \left.\dfrac{e^{i(k-j)\omega}}{i(k-j)}\right|_0^{2\pi} = 0 & k \neq j \\ 2\pi & k = j \end{cases}$$

Therefore,

$$p_j = \frac{1}{2\pi}\int_{\omega=0}^{2\pi} e^{-ij\omega}\phi(\omega)\,d\omega$$

or, replacing j by k,

$$p_k = \frac{1}{2\pi}\int_{\omega=0}^{2\pi} e^{-ik\omega}\phi(\omega)\,d\omega$$

We often call $\sum_{k=-n}^{n} p_k e^{ik\omega}$ (where n can theoretically be infinite) the *Fourier series* of $\phi(\omega)$ and p_k the *Fourier coefficients.* For a continuous random variable, the Fourier series is replaced by the Fourier integral (see page 81).

3.40. Use Problem 3.39 to obtain the probability distribution of a random variable X whose characteristic function is $\phi(\omega) = \cos\omega$.

From Problem 3.39

$$\begin{aligned} p_k &= \frac{1}{2\pi}\int_{\omega=0}^{2\pi} e^{-ik\omega}\cos\omega\,d\omega \\ &= \frac{1}{2\pi}\int_{\omega=0}^{2\pi} e^{-ik\omega}\left[\frac{e^{i\omega}+e^{-i\omega}}{2}\right]d\omega \\ &= \frac{1}{4\pi}\int_{\omega=0}^{2\pi} e^{i(1-k)\omega}\,d\omega + \frac{1}{4\pi}\int_{\omega=0}^{2\pi} e^{-i(1+k)\omega}\,d\omega \end{aligned}$$

If $k = 1$, we find $p_1 = \frac{1}{2}$; if $k = -1$, we find $p_{-1} = \frac{1}{2}$. For all other values of k, we have $p_k = 0$. Therefore, the random variable is given by

$$X = \begin{cases} 1 & \text{probability } 1/2 \\ -1 & \text{probability } 1/2 \end{cases}$$

As a check, see Problem 3.20.

3.41. Find the coefficient of (a) skewness, (b) kurtosis of the distribution defined by the *normal curve*, having density

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \qquad -\infty < x < \infty$$

(a) The distribution has the appearance of Fig. 3-7. By symmetry, $\mu_1' = \mu = 0$ and $\mu_3' = 0$. Therefore the coefficient of skewness is zero.

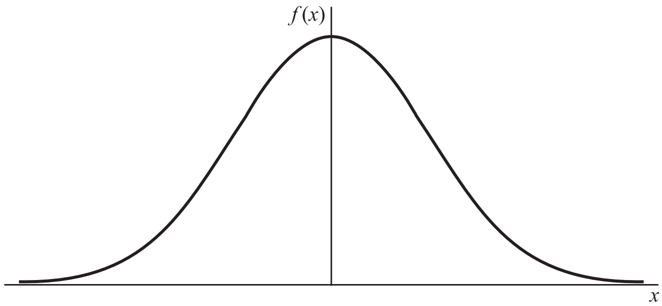


Fig. 3-7

(b) We have

$$\mu_2' = E(X^2) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2 e^{-x^2/2}\,dx = \frac{2}{\sqrt{2\pi}}\int_0^{\infty} x^2 e^{-x^2/2}\,dx$$
$$= \frac{2}{\sqrt{\pi}}\int_0^{\infty} v^{1/2}e^{-v}\,dv$$
$$= \frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}\right) = \frac{2}{\sqrt{\pi}}\cdot\frac{1}{2}\Gamma\left(\frac{1}{2}\right) = 1$$

where we have made the transformation $x^2/2 = v$ and used properties of the gamma function given in (2) and (5) of Appendix A. Similarly we obtain

$$\mu_4' = E(X^4) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^4 e^{-x^2/2}\,dx = \frac{2}{\sqrt{2\pi}}\int_0^{\infty} x^4 e^{-x^2/2}\,dx$$
$$= \frac{4}{\sqrt{\pi}}\int_0^{\infty} v^{3/2}e^{-v}\,dv$$
$$= \frac{4}{\sqrt{\pi}}\Gamma\left(\frac{5}{2}\right) = \frac{4}{\sqrt{\pi}}\cdot\frac{3}{2}\cdot\frac{1}{2}\Gamma\left(\frac{1}{2}\right) = 3$$

Now

$$\sigma^2 = E[(X - \mu)^2] = E(X)^2 = \mu_2' = 1$$
$$\mu_4 = E[(X - \mu)^4] = E(X^4) = \mu_4' = 3$$

Thus the coefficient of kurtosis is

$$\frac{\mu_4}{\sigma^4} = 3$$

3.42. Prove that $-1 \le \rho \le 1$ (see page 82).

For any real constant c, we have

$$E[\{Y - \mu_Y - c(X - \mu)\}^2] \ge 0$$

Now the left side can be written

$$E[(Y - \mu_Y)^2] + c^2E[(X - \mu_X)^2] - 2cE[(X - \mu_X)(Y - \mu_Y)] = \sigma_Y^2 + c^2\sigma_X^2 - 2c\sigma_{XY}$$
$$= \sigma_Y^2 + \sigma_X^2\left(c^2 - \frac{2c\sigma_{XY}}{\sigma_X^2}\right)$$
$$= \sigma_Y^2 + \sigma_X^2\left(c^2 - \frac{\sigma_{XY}}{\sigma_X^2}\right)^2 - \frac{\sigma_{XY}^2}{\sigma_X^2}$$
$$= \frac{\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2}{\sigma_X^2} + \sigma_X^2\left(c - \frac{\sigma_{XY}}{\sigma_X^2}\right)^2$$

In order for this last quantity to be greater than or equal to zero for every value of c, we must have

$$\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2 \geq 0 \quad \text{or} \quad \frac{\sigma_{XY}^2}{\sigma_X^2\sigma_Y^2} \leq 1$$

which is equivalent to $\rho^2 \leq 1$ or $-1 \leq \rho \leq 1$.

SUPPLEMENTARY PROBLEMS

Expectation of random variables

3.43. A random variable X is defined by $X = \begin{cases} -2 & \text{prob. } 1/3 \\ 3 & \text{prob. } 1/2. \\ 1 & \text{prob. } 1/6 \end{cases}$ Find (a) $E(X)$, (b) $E(2X + 5)$, (c) $E(X^2)$.

3.44. Let X be a random variable defined by the density function $f(x) = \begin{cases} 3x^2 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$.

Find (a) $E(X)$, (b) $E(3X - 2)$, (c) $E(X^2)$.

3.45. The density function of a random variable X is $f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$.

Find (a) $E(X)$, (b) $E(X^2)$, (c) $E[(X - 1)^2]$.

3.46. What is the expected number of points that will come up in 3 successive tosses of a fair die? Does your answer seem reasonable? Explain.

3.47. A random variable X has the density function $f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$. Find $E(e^{2X/3})$.

3.48. Let X and Y be independent random variables each having density function

$$f(u) = \begin{cases} 2e^{-2u} & u \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) $E(X + Y)$, (b) $E(X^2 + Y^2)$, (c) $E(XY)$.

3.49. Does (a) $E(X + Y) = E(X) + E(Y)$, (b) $E(XY) = E(X)E(Y)$, in Problem 3.48? Explain.

3.50. Let X and Y be random variables having joint density function

$$f(x, y) = \begin{cases} \frac{3}{5}x(x + y) & 0 \leq x \leq 1, 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) $E(X)$, (b) $E(Y)$, (c) $E(X + Y)$, (d) $E(XY)$.

3.51. Does (a) $E(X + Y) = E(X) + E(Y)$, (b) $E(XY) = E(X)E(Y)$, in Problem 3.50? Explain.

3.52. Let X and Y be random variables having joint density

$$f(x, y) = \begin{cases} 4xy & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) $E(X)$, (b) $E(Y)$, (c) $E(X + Y)$, (d) $E(XY)$.

3.53. Does (a) $E(X + Y) = E(X) + E(Y)$, (b) $E(XY) = E(X)E(Y)$, in Problem 3.52? Explain.

3.54. Let $f(x, y) = \begin{cases} \frac{1}{4}(2x + y) & 0 \le x \le 1, 0 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}$. Find (a) $E(X)$, (b) $E(Y)$, (c) $E(X^2)$, (d) $E(Y^2)$, (e) $E(X + Y)$, (f) $E(XY)$.

3.55. Let X and Y be independent random variables such that

$$X = \begin{cases} 1 & \text{prob. } 1/3 \\ 0 & \text{prob. } 2/3 \end{cases} \qquad Y = \begin{cases} 2 & \text{prob. } 3/4 \\ -3 & \text{prob. } 1/4 \end{cases}$$

Find (a) $E(3X + 2Y)$, (b) $E(2X^2 - Y^2)$, (c) $E(XY)$, (d) $E(X^2Y)$.

3.56. Let $X_1, X_2, \ldots, X_n$ be n random variables which are identically distributed such that

$$X_k = \begin{cases} 1 & \text{prob. } 1/2 \\ 2 & \text{prob. } 1/3 \\ -1 & \text{prob. } 1/6 \end{cases}$$

Find (a) $E(X_1 + X_2 + \cdots + X_n)$, (b) $E(X_1^2 + X_2^2 + \cdots + X_n^2)$.

Variance and standard deviation

3.57. Find (a) the variance, (b) the standard deviation of the number of points that will come up on a single toss of a fair die.

3.58. Let X be a random variable having density function

$$f(x) = \begin{cases} 1/4 & -2 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) $\text{Var}(X)$, (b) σ_X.

3.59. Let X be a random variable having density function

$$f(x) = \begin{cases} e^{-x} & x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) $\text{Var}(X)$, (b) σ_X.

3.60. Find the variance and standard deviation for the random variable X of (a) Problem 3.43, (b) Problem 3.44.

3.61. A random variable X has $E(X) = 2$, $E(X^2) = 8$. Find (a) $\text{Var}(X)$, (b) σ_X.

3.62. If a random variable X is such that $E[(X - 1)^2] = 10$, $E[(X - 2)^2] = 6$ find (a) $E(X)$, (b) $\text{Var}(X)$, (c) σ_X.

Moments and moment generating functions

3.63. Find (a) the moment generating function of the random variable

$$X = \begin{cases} 1/2 & \text{prob. } 1/2 \\ -1/2 & \text{prob. } 1/2 \end{cases}$$

and (b) the first four moments about the origin.

3.64. (a) Find the moment generating function of a random variable X having density function

$$f(x) = \begin{cases} x/2 & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

(b) Use the generating function of (a) to find the first four moments about the origin.

3.65. Find the first four moments about the mean in (a) Problem 3.43, (b) Problem 3.44.

3.66. (a) Find the moment generating function of a random variable having density function

$$f(x) = \begin{cases} e^{-x} & x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

and (b) determine the first four moments about the origin.

3.67. In Problem 3.66 find the first four moments about the mean.

3.68. Let X have density function $f(x) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$. Find the kth moment about (a) the origin, (b) the mean.

3.69. If $M(t)$ is the moment generating function of the random variable X, prove that the 3rd and 4th moments about the mean are given by

$$\mu_3 = M'''(0) - 3M''(0)M'(0) + 2[M'(0)]^3$$
$$\mu_4 = M^{(\text{iv})}(0) - 4M'''(0)M'(0) + 6M''(0)[M'(0)]^2 - 3[M'(0)]^4$$

Characteristic functions

3.70. Find the characteristic function of the random variable $X = \begin{cases} a & \text{prob. } p \\ b & \text{prob. } q = 1 - p \end{cases}$.

3.71. Find the characteristic function of a random variable X that has density function

$$f(x) = \begin{cases} 1/2a & |x| \le a \\ 0 & \text{otherwise} \end{cases}$$

3.72. Find the characteristic function of a random variable with density function

$$f(x) = \begin{cases} x/2 & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

3.73. Let $X_k = \begin{cases} 1 & \text{prob. } 1/2 \\ -1 & \text{prob. } 1/2 \end{cases}$ be independent random variables ($k = 1, 2, \ldots, n$). Prove that the characteristic function of the random variable

$$\frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}$$

is $[\cos(\omega/\sqrt{n})]^n$.

3.74. Prove that as $n \to \infty$ the characteristic function of Problem 3.73 approaches $e^{-\omega^2/2}$. (*Hint*: Take the logarithm of the characteristic function and use L'Hospital's rule.)

Covariance and correlation coefficient

3.75. Let X and Y be random variables having joint density function

$$f(x, y) = \begin{cases} x + y & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) $\mathrm{Var}(X)$, (b) $\mathrm{Var}(Y)$, (c) σ_X, (d) σ_Y, (e) σ_{XY}, (f) ρ.

3.76. Work Problem 3.75 if the joint density function is $f(x, y) = \begin{cases} e^{-(x+y)} & x \ge 0,\ y \ge 0 \\ 0 & \text{otherwise} \end{cases}$.

3.77. Find (a) $\mathrm{Var}(X)$, (b) $\mathrm{Var}(Y)$, (c) σ_X, (d) σ_Y, (e) σ_{XY}, (f) ρ, for the random variables of Problem 2.56.

3.78. Work Problem 3.77 for the random variables of Problem 2.94.

3.79. Find (a) the covariance, (b) the correlation coefficient of two random variables X and Y if $E(X) = 2$, $E(Y) = 3$, $E(XY) = 10$, $E(X^2) = 9$, $E(Y^2) = 16$.

3.80. The correlation coefficient of two random variables X and Y is $-\frac{1}{4}$ while their variances are 3 and 5. Find the covariance.

Conditional expectation, variance, and moments

3.81. Let X and Y have joint density function

$$f(x, y) = \begin{cases} x + y & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the conditional expectation of (a) Y given X, (b) X given Y.

3.82. Work Problem 3.81 if $f(x, y) = \begin{cases} 2e^{-(x+2y)} & x \ge 0,\ y \ge 0 \\ 0 & \text{otherwise} \end{cases}$

3.83. Let X and Y have the joint probability function given in Table 2-9, page 71. Find the conditional expectation of (a) Y given X, (b) X given Y.

3.84. Find the conditional variance of (a) Y given X, (b) X given Y for the distribution of Problem 3.81.

3.85. Work Problem 3.84 for the distribution of Problem 3.82.

3.86. Work Problem 3.84 for the distribution of Problem 2.94.

Chebyshev's inequality

3.87. A random variable X has mean 3 and variance 2. Use Chebyshev's inequality to obtain an upper bound for (a) $P(|X - 3| \ge 2)$, (b) $P(|X - 3| \ge 1)$.

3.88. Prove Chebyshev's inequality for a discrete variable X. *(Hint*: See Problem 3.30.)

3.89. A random variable X has the density function $f(x) = \frac{1}{2}e^{-|x|}$, $-\infty < x < \infty$. (a) Find $P(|X - \mu| > 2)$. (b) Use Chebyshev's inequality to obtain an upper bound on $P(|X - \mu| > 2)$ and compare with the result in (a).

Law of large numbers

3.90. Show that the (weak) law of large numbers can be stated as

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| < \epsilon\right) = 1$$

and interpret.

3.91. Let X_k ($k = 1, \ldots, n$) be n independent random variables such that

$$X_k = \begin{cases} 1 & \text{prob. } p \\ 0 & \text{prob. } q = 1 - p \end{cases}$$

(a) If we interpret X_k to be the number of heads on the kth toss of a coin, what interpretation can be given to $S_n = X_1 + \cdots + X_n$?

(b) Show that the law of large numbers in this case reduces to

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) = 0$$

and interpret this result.

Other measures of central tendency

3.92. Find (a) the mode, (b) the median of a random variable X having density function

$$f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

and (c) compare with the mean.

3.93. Work Problem 3.100 if the density function is

$$f(x) = \begin{cases} 4x(1 - x^2) & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

3.94. Find (a) the median, (b) the mode for a random variable X defined by

$$X = \begin{cases} 2 & \text{prob. } 1/3 \\ -1 & \text{prob. } 2/3 \end{cases}$$

and (c) compare with the mean.

3.95. Find (a) the median, (b) the mode of the set of numbers 1, 3, 2, 1, 5, 6, 3, 3, and (c) compare with the mean.

Percentiles

3.96. Find the (a) 25th, (b) 75th percentile values for the random variable having density function

$$f(x) = \begin{cases} 2(1 - x) & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

3.97. Find the (a) 10th, (b) 25th, (c) 75th, (d) 90th percentile values for the random variable having density function

$$f(x) = \begin{cases} c(x - x^3) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where c is an appropriate constant.

Other measures of dispersion

3.98. Find (a) the semi-interquartile range, (b) the mean deviation for the random variable of Problem 3.96.

3.99. Work Problem 3.98 for the random variable of Problem 3.97.

3.100. Find the mean deviation of the random variable X in each of the following cases.

(a) $f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$ (b) $f(x) = \dfrac{1}{\pi(1 + x^2)}, \quad -\infty < x < \infty.$

3.101. Obtain the probability that the random variable X differs from its mean by more than the semi-interquartile range in the case of (a) Problem 3.96, (b) Problem 3.100(a).

Skewness and kurtosis

3.102. Find the coefficient of (a) skewness, (b) kurtosis for the distribution of Problem 3.100(a).

3.103. If

$$f(x) = \begin{cases} c\left(1 - \dfrac{|x|}{a}\right) & |x| \leq a \\ 0 & |x| > a \end{cases}$$

where c is an appropriate constant, is the density function of X, find the coefficient of (a) skewness, (b) kurtosis.

3.104. Find the coefficient of (a) skewness, (b) kurtosis, for the distribution with density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Miscellaneous problems

3.105. Let X be a random variable that can take on the values 2, 1, and 3 with respective probabilities $1/3$, $1/6$, and $1/2$. Find (a) the mean, (b) the variance, (c) the moment generating function, (d) the characteristic function, (e) the third moment about the mean.

3.106. Work Problem 3.105 if X has density function

$$f(x) = \begin{cases} c(1 - x) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where c is an appropriate constant.

3.107. Three dice, assumed fair, are tossed successively. Find (a) the mean, (b) the variance of the sum.

3.108. Let X be a random variable having density function

$$f(x) = \begin{cases} cx & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

where c is an appropriate constant. Find (a) the mean, (b) the variance, (c) the moment generating function, (d) the characteristic function, (e) the coefficient of skewness, (f) the coefficient of kurtosis.

3.109. Let X and Y have joint density function

$$f(x, y) = \begin{cases} cxy & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) $E(X^2 + Y^2)$, (b) $E(\sqrt{X^2 + Y^2})$.

3.110. Work Problem 3.109 if X and Y are independent identically distributed random variables having density function $f(u) = (2\pi)^{-1/2}e^{-u^2/2}$, $-\infty < u < \infty$.

3.111. Let X be a random variable having density function

$$f(x) = \begin{cases} \frac{1}{2} & -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

and let $Y = X^2$. Find (a) $E(X)$, (b) $E(Y)$, (c) $E(XY)$.

ANSWERS TO SUPPLEMENTARY PROBLEMS

3.43. (a) 1 (b) 7 (c) 6 **3.44.** (a) $3/4$ (b) $1/4$ (c) $3/5$

3.45. (a) 1 (b) 2 (c) 1 **3.46.** 10.5 **3.47.** 3

3.48. (a) 1 (b) 1 (c) $1/4$

3.50. (a) $7/10$ (b) $6/5$ (c) $19/10$ (d) $5/6$

3.52. (a) $2/3$ (b) $2/3$ (c) $4/3$ (d) $4/9$

3.54. (a) $7/12$ (b) $7/6$ (c) $5/12$ (d) $5/3$ (e) $7/4$ (f) $2/3$

3.55. (a) $5/2$ (b) $-55/12$ (c) $1/4$ (d) $1/4$

3.56. (a) n (b) $2n$ **3.57.** (a) $35/12$ (b) $\sqrt{35/12}$

3.58. (a) $4/3$ (b) $\sqrt{4/3}$ **3.59.** (a) 1 (b) 1

3.60. (a) $\text{Var}(X) = 5, \sigma_X = \sqrt{5}$ (b) $\text{Var}(X) = 3/80, \sigma_X = \sqrt{15}/20$

3.61. (a) 4 (b) 2 **3.62.** (a) $7/2$ (b) $15/4$ (c) $\sqrt{15}/2$

3.63. (a) $\frac{1}{2}(e^{t/2} + e^{-t/2}) = \cosh(t/2)$ (b) $\mu = 0, \mu_2' = 1, \mu_3' = 0, \mu_4' = 1$

3.64. (a) $(1 + 2te^{2t} - e^{2t})/2t^2$ (b) $\mu = 4/3, \mu_2' = 2, \mu_3' = 16/5, \mu_4' = 16/3$

3.65. (a) $\mu_1 = 0, \mu_2 = 5, \mu_3 = -5, \mu_4 = 35$ (b) $\mu_1 = 0, \mu_2 = 3/80, \mu_3 = -121/160, \mu_4 = 2307/8960$

3.66. (a) $1/(1 - t), |t| < 1$ (b) $\mu = 1, \mu_2' = 2, \mu_3' = 6, \mu_4' = 24$

3.67. $\mu_1 = 0, \mu_2 = 1, \mu_3 = 2, \mu_4 = 33$

3.68. (a) $(b^{k+1} - a^{k+1})/(k + 1)(b - a)$ (b) $[1 + (-1)^k](b - a)^k/2^{k+1}(k + 1)$

3.70. $pe^{i\omega a} + qe^{i\omega b}$ **3.71.** $(\sin a\omega)/a\omega$ **3.72.** $(e^{2i\omega} - 2i\omega e^{2i\omega} - 1)/2\omega^2$

3.75. (a) $11/144$ (b) $11/144$ (c) $\sqrt{11}/12$ (d) $\sqrt{11}/12$ (e) $-1/144$ (f) $-1/11$

3.76. (a) 1 (b) 1 (c) 1 (d) 1 (e) 0 (f) 0

3.77. (a) $73/960$ (b) $73/960$ (c) $\sqrt{73/960}$ (d) $\sqrt{73/960}$ (e) $-1/64$ (f) $-15/73$

3.78. (a) $233/324$ (b) $233/324$ (c) $\sqrt{233}/18$ (d) $\sqrt{233}/18$ (e) $-91/324$ (f) $-91/233$

3.79. (a) 4 (b) $4/\sqrt{35}$ **3.80.** $-\sqrt{15}/4$

3.81. (a) $(3x + 2)/(6x + 3)$ for $0 \le x \le 1$ (b) $(3y + 2)/(6y + 3)$ for $0 \le y \le 1$

3.82. (a) $1/2$ for $x \ge 0$ (b) 1 for $y \ge 0$

3.83. (a)

X	0	1	2
$E(Y\,\vert\,X)$	4/3	1	5/7

(b)

Y	0	1	2
$E(X\,\vert\,Y)$	4/3	7/6	1/2

3.84. (a) $\dfrac{6x^2 + 6x + 1}{18(2x + 1)^2}$ for $0 \le x \le 1$ (b) $\dfrac{6y^2 + 6y + 1}{18(2y + 1)^2}$ for $0 \le y \le 1$

3.85. (a) $1/9$ (b) 1

3.86. (a)

X	0	1	2
$\mathrm{Var}(Y\,\vert\,X)$	5/9	4/5	24/49

(b)

Y	0	1	2
$\mathrm{Var}(X\,\vert\,Y)$	5/9	29/36	7/12

3.87. (a) $1/2$ (b) 2 (useless) **3.89.** (a) e^{-2} (b) 0.5

3.92. (a) $+0$ (b) $\ln 2$ (c) 1 **3.93.** (a) $1/\sqrt{3}$ (b) $\sqrt{1 - (1/\sqrt{2})}$ (c) $8/15$

3.94. (a) does not exist (b) -1 (c) 0 **3.95.** (a) 3 (b) 3 (c) 3

3.96. (a) $1 - \frac{1}{2}\sqrt{3}$ (b) $1/2$

3.97. (a) $\sqrt{1 - (3/\sqrt{10})}$ (b) $\sqrt{1 - (\sqrt{3}/2)}$ (c) $\sqrt{1/2}$ (d) $\sqrt{1 - (1/\sqrt{10})}$

3.98. (a) 1 (b) $(\sqrt{3} - 1)/4$ (c) $16/81$

3.99. (a) 1 (b) 0.17 (c) 0.051 **3.100.** (a) $1 - 2e^{-1}$ (b) does not exist

3.101. (a) $(5 - 2\sqrt{3})/3$ (b) $(3 - 2e^{-1}\sqrt{3})/3$

3.102. (a) 2 (b) 9 **3.103.** (a) 0 (b) $24/5a$ **3.104.** (a) 2 (b) 9

3.105. (a) $7/3$ (b) $5/9$ (c) $(e^t + 2e^{2t} + 3e^{3t})/6$ (d) $(e^{i\omega} + 2e^{2i\omega} + 3e^{3i\omega})/6$ (e) $-7/27$

3.106. (a) $1/3$ (b) $1/18$ (c) $2(e^t - 1 - t)/t^2$ (d) $-2(e^{i\omega} - 1 - i\omega)/\omega^2$ (*e*) $1/135$

3.107. (a) $21/2$ (b) $35/4$

3.108. (a) $4/3$ (b) $2/9$ (c) $(1 + 2te^{2t} - e^{2t})/2t^2$ (d) $-(1 + 2i\omega e^{2i\omega} - e^{2i\omega})/2\omega^2$
(e) $-2\sqrt{18}/15$ (f) $12/5$

3.109. (a) 1 (b) $8(2\sqrt{2} - 1)/15$

3.110. (a) 2 (b) $\sqrt{2\pi}/2$

3.111. (a) 0 (b) $1/3$ (c) 0

CHAPTER 4

Special Probability Distributions

The Binomial Distribution

Suppose that we have an experiment such as tossing a coin or die repeatedly or choosing a marble from an urn repeatedly. Each toss or selection is called a *trial*. In any single trial there will be a probability associated with a particular event such as head on the coin, 4 on the die, or selection of a red marble. In some cases this probability will not change from one trial to the next (as in tossing a coin or die). Such trials are then said to be *independent* and are often called *Bernoulli trials* after James Bernoulli who investigated them at the end of the seventeenth century.

Let p be the probability that an event will happen in any single Bernoulli trial (called the *probability of success*). Then $q = 1 - p$ is the probability that the event will fail to happen in any single trial (called the *probability of failure*). The probability that the event will happen exactly x times in n trials (i.e., successes and $n - x$ failures will occur) is given by the probability function

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x} \tag{1}$$

where the random variable X denotes the number of successes in n trials and $x = 0, 1, \dots, n$.

EXAMPLE 4.1 The probability of getting exactly 2 heads in 6 tosses of a fair coin is

$$P(X = 2) = \binom{6}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{6-2} = \frac{6!}{2!4!}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{6-2} = \frac{15}{64}$$

The discrete probability function (1) is often called the *binomial distribution* since for $x = 0, 1, 2, \dots, n$, it corresponds to successive terms in the *binomial expansion*

$$(q + p)^n = q^n + \binom{n}{1}q^{n-1}p + \binom{n}{2}q^{n-2}p^2 + \cdots + p^n = \sum_{x=0}^{n}\binom{n}{x}p^x q^{n-x} \tag{2}$$

The special case of a binomial distribution with $n = 1$ is also called the *Bernoulli distribution*.

Some Properties of the Binomial Distribution

Some of the important properties of the binomial distribution are listed in Table 4-1.

Table 4-1

Mean	$\mu = np$
Variance	$\sigma^2 = npq$
Standard deviation	$\sigma = \sqrt{npq}$
Coefficient of skewness	$\alpha_3 = \dfrac{q - p}{\sqrt{npq}}$
Coefficient of kurtosis	$\alpha_4 = 3 + \dfrac{1 - 6pq}{npq}$
Moment generating function	$M(t) = (q + pe^t)^n$
Characteristic function	$\phi(\omega) = (q + pe^{i\omega})^n$

EXAMPLE 4.2 In 100 tosses of a fair coin, the expected or mean number of heads is $\mu = (100)\left(\frac{1}{2}\right) = 50$ while the standard deviation is $\sigma = \sqrt{(100)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = 5$.

The Law of Large Numbers for Bernoulli Trials

The law of large numbers, page 83, has an interesting interpretation in the case of Bernoulli trials and is presented in the following theorem.

Theorem 4-1 **(Law of Large Numbers for Bernoulli Trials):** Let X be the random variable giving the number of successes in n Bernoulli trials, so that X/n is the proportion of successes. Then if p is the probability of success and ϵ is any positive number,

$$\lim_{n\to\infty} P\left(\left|\frac{X}{n} - p\right| \geq \epsilon\right) = 0 \tag{3}$$

In other words, in the long run it becomes extremely likely that the proportion of successes, X/n, will be as close as you like to the probability of success in a single trial, p. This law in a sense justifies use of the empirical definition of probability on page 5. A stronger result is provided by the *strong law* of large numbers (page 83), which states that with probability one, $\lim_{n\to\infty} X/n = p$, i.e., X/n actually *converges to p* except in a negligible number of cases.

The Normal Distribution

One of the most important examples of a continuous probability distribution is the *normal distribution*, sometimes called the *Gaussian distribution*. The density function for this distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty \tag{4}$$

where μ and σ are the mean and standard deviation, respectively. The corresponding distribution function is given by

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(v-\mu)^2/2\sigma^2}\, dv \tag{5}$$

If X has the distribution function given by (5), we say that the random variable X is *normally distributed* with mean μ and variance σ^2.

If we let Z be the standardized variable corresponding to X, i.e., if we let

$$Z = \frac{X - \mu}{\sigma} \tag{6}$$

then the mean or expected value of Z is 0 and the variance is 1. In such cases the density function for Z can be obtained from (4) by formally placing $\mu = 0$ and $\sigma = 1$, yielding

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tag{7}$$

This is often referred to as the *standard normal density function.* The corresponding distribution function is given by

$$F(z) = P(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\,du = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_{0}^{z} e^{-u^2/2}\,du \tag{8}$$

We sometimes call the value z of the standardized variable Z the *standard score.* The function $F(z)$ is related to the extensively tabulated *error function*, erf(z). We have

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-u^2}\,du \qquad \text{and} \qquad F(z) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)\right] \tag{9}$$

A graph of the density function (7), sometimes called the *standard normal curve*, is shown in Fig. 4-1. In this graph we have indicated the areas within 1, 2, and 3 standard deviations of the mean (i.e., between $z = -1$ and $+1$, $z = -2$ and $+2$, $z = -3$ and $+3$) as equal, respectively, to 68.27%, 95.45% and 99.73% of the total area, which is one. This means that

$$P(-1 \le Z \le 1) = 0.6827, \qquad P(-2 \le Z \le 2) = 0.9545, \qquad P(-3 \le Z \le 3) = 0.9973 \tag{10}$$

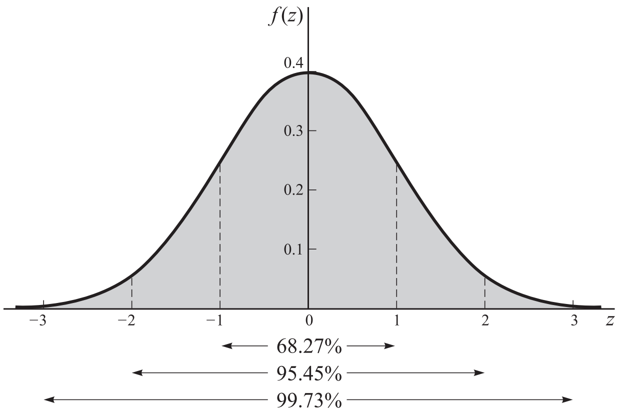


Fig. 4-1

A table giving the areas under this curve bounded by the ordinates at $z = 0$ and any positive value of z is given in Appendix C. From this table the areas between any two ordinates can be found by using the symmetry of the curve about $z = 0$.

Some Properties of the Normal Distribution

In Table 4-2 we list some important properties of the general normal distribution.

Table 4-2

Mean	μ
Variance	σ^2
Standard deviation	σ
Coefficient of skewness	$\alpha_3 = 0$
Coefficient of kurtosis	$\alpha_4 = 3$
Moment generating function	$M(t) = e^{ut+(\sigma^2 t^2/2)}$
Characteristic function	$\phi(\omega) = e^{i\mu\omega-(\sigma^2\omega^2/2)}$

Relation Between Binomial and Normal Distributions

If n is large and if neither p nor q is too close to zero, the binomial distribution can be closely approximated by a normal distribution with standardized random variable given by

$$Z = \frac{X - np}{\sqrt{npq}} \tag{11}$$

Here X is the random variable giving the number of successes in n Bernoulli trials and p is the probability of success. The approximation becomes better with increasing n and is exact in the limiting case. (See Problem 4.17.) In practice, the approximation is very good if both np and nq are greater than 5. The fact that the binomial distribution approaches the normal distribution can be described by writing

$$\lim_{n \to \infty} P\left(a \leq \frac{X - np}{\sqrt{npq}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-u^2/2}\, du \tag{12}$$

In words, we say that the standardized random variable $(X - np)/\sqrt{npq}$ is *asymptotically normal*.

The Poisson Distribution

Let X be a discrete random variable that can take on the values 0, 1, 2, . . . such that the probability function of X is given by

$$f(x) = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0, 1, 2, \ldots \tag{13}$$

where λ is a given positive constant. This distribution is called the *Poisson distribution* (after S. D. Poisson, who discovered it in the early part of the nineteenth century), and a random variable having this distribution is said to be *Poisson distributed*.

The values of $f(x)$ in (13) can be obtained by using Appendix G, which gives values of $e^{-\lambda}$ for various values of λ.

Some Properties of the Poisson Distribution

Some important properties of the Poisson distribution are listed in Table 4-3.

Table 4-3

Mean	$\mu = \lambda$
Variance	$\sigma^2 = \lambda$
Standard deviation	$\sigma = \sqrt{\lambda}$
Coefficient of skewness	$\alpha_3 = 1/\sqrt{\lambda}$
Coefficient of kurtosis	$\alpha_4 = 3 + (1/\lambda)$
Moment generating function	$M(t) = e^{\lambda(e^t - 1)}$
Characteristic function	$\phi(\omega) = e^{\lambda(e^{i\omega} - 1)}$

Relation Between the Binomial and Poisson Distributions

In the binomial distribution (1), if n is large while the probability p of occurrence of an event is close to zero, so that $q = 1 - p$ is close to 1, the event is called a *rare event*. In practice we shall consider an event as rare if the number of trials is at least 50 ($n \geq 50$) while np is less than 5. For such cases the binomial distribution is very closely approximated by the Poisson distribution (13) with $\lambda = np$. This is to be expected on comparing Tables 4-1 and 4-3, since by placing $\lambda = np$, $q \approx 1$, and $p \approx 0$ in Table 4-1, we get the results in Table 4-3.

Relation Between the Poisson and Normal Distributions

Since there is a relation between the binomial and normal distributions and between the binomial and Poisson distributions, we would expect that there should also be a relation between the Poisson and normal distributions. This is in fact the case. We can show that if X is the Poisson random variable of (13) and $(X - \lambda)/\sqrt{\lambda}$ is the corresponding standardized random variable, then

$$\lim_{\lambda \to \infty} P\left(a \leq \frac{X - \lambda}{\sqrt{\lambda}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-u^2/2}\, du \tag{14}$$

i.e., the Poisson distribution approaches the normal distribution as $\lambda \to \infty$ or $(X - \lambda)/\sqrt{\lambda}$ is *asymptotically normal.*

The Central Limit Theorem

The similarity between (12) and (14) naturally leads us to ask whether there are any other distributions besides the binomial and Poisson that have the normal distribution as the limiting case. The following remarkable theorem reveals that actually a large class of distributions have this property.

Theorem 4-2 **(Central Limit Theorem)** Let $X_1, X_2, \ldots, X_n$ be independent random variables that are identically distributed (i.e., all have the *same* probability function in the discrete case or density function in the continuous case) and have finite mean μ and variance σ^2. Then if $S_n = X_1 + X_2 + \cdots + X_n$ $(n = 1, 2 \ldots)$,

$$\lim_{n \to \infty} P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-u^2/2}\, du \tag{15}$$

that is, the random variable $(S_n - n\mu)/\sigma\sqrt{n}$, which is the standardized variable corresponding to S_n, is asymptotically normal.

The theorem is also true under more general conditions; for example, it holds when $X_1, X_2, \ldots, X_n$ are independent random variables with the same mean and the same variance but not necessarily identically distributed.

The Multinomial Distribution

Suppose that events $A_1, A_2, \ldots, A_k$ are mutually exclusive, and can occur with respective probabilities $p_1, p_2, \ldots, p_k$ where $p_1 + p_2 + \cdots + p_k = 1$. If $X_1, X_2, \ldots, X_k$ are the random variables respectively giving the number of times that $A_1, A_2, \ldots, A_k$ occur in a total of n trials, so that $X_1 + X_2 + \cdots + X_k = n$, then

$$P(X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k) = \frac{n}{n_1!n_2! \cdots n_k!} p_1^{n_1} p_k^{n_2} \cdots p_k^{n_k} \tag{16}$$

where $n_1 + n_2 + \cdots + n_k = n$, is the joint probability function for the random variables $X_1, \ldots, X_k$.

This distribution, which is a generalization of the binomial distribution, is called the *multinomial distribution* since (16) is the general term in the multinomial expansion of $(p_1 + p_2 + \cdots + p_k)^n$.

EXAMPLE 4.3 If a fair die is to be tossed 12 times, the probability of getting 1, 2, 3, 4, 5 and 6 points exactly twice each is

$$P(X_1 = 2, X_2 = 2, \ldots, X_6 = 2) = \frac{12!}{2!2!2!2!2!2!}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2 = \frac{1925}{559{,}872} = 0.00344$$

The expected number of times that $A_1, A_2, \ldots, A_k$ will occur in n trials are $np_1, np_2, \ldots, np_k$ respectively, i.e.,

$$E(X_1) = np_1, \qquad E(X_2) = np_2, \qquad \ldots, \qquad E(X_k) = np_k \tag{17}$$

The Hypergeometric Distribution

Suppose that a box contains b blue marbles and r red marbles. Let us perform n trials of an experiment in which a marble is chosen at random, its color is observed, and then the marble is put back in the box. This type of experiment is often referred to as *sampling with replacement.* In such a case, if X is the random variable denoting the number of blue marbles chosen (successes) in n trials, then using the binomial distribution (1) we see that the probability of exactly x successes is

$$P(X = x) = \binom{n}{x} \frac{b^x r^{n-x}}{(b + r)^n}, \qquad x = 0, 1, \ldots, n \tag{18}$$

since $p = b/(b + r)$, $q = 1 - p = r/(b + r)$.

If we modify the above so that *sampling is without replacement*, i.e., the marbles are not replaced after being chosen, then

$$P(X = x) = \frac{\binom{b}{x}\binom{r}{n-x}}{\binom{b+r}{n}}, \qquad x = \max(0, n - r), \ldots, \min(n, b) \tag{19}$$

This is the *hypergeometric distribution.* The mean and variance for this distribution are

$$\mu = \frac{nb}{b + r}, \qquad \sigma^2 = \frac{nbr(b + r - n)}{(b + r)^2 (b + r - 1)} \tag{20}$$

If we let the total number of blue and red marbles be N, while the proportions of blue and red marbles are p and $q = 1 - p$, respectively, then

$$p = \frac{b}{b + r} = \frac{b}{N}, \qquad q = \frac{r}{b + r} = \frac{r}{N} \qquad \text{or} \qquad b = Np, \qquad r = Nq \tag{21}$$

so that (19) and (20) become, respectively,

$$P(X = x) = \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}} \tag{22}$$

$$\mu = np, \qquad \sigma^2 = \frac{npq(N - n)}{N - 1} \tag{23}$$

Note that as $N \to \infty$ (or N is large compared with n), (22) reduces to (18), which can be written

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \tag{24}$$

and (23) reduces to

$$\mu = np, \qquad \sigma^2 = npq \tag{25}$$

in agreement with the first two entries in Table 4-1, page 109. The results are just what we would expect, since for large N, sampling without replacement is practically identical to sampling with replacement.

The Uniform Distribution

A random variable X is said to be *uniformly distributed* in $a \le x \le b$ if its density function is

$$f(x) = \begin{cases} 1/(b - a) & a \le x \le b \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

and the distribution is called a *uniform distribution*.

The distribution function is given by

$$F(x) = P(X \le x) = \begin{cases} 0 & x < a \\ (x - a)/(b - a) & a \le x < b \\ 1 & x \ge b \end{cases} \tag{27}$$

The mean and variance are, respectively,

$$\mu = \frac{1}{2}(a + b), \qquad \sigma^2 = \frac{1}{12}(b - a)^2 \tag{28}$$

The Cauchy Distribution

A random variable X is said to be *Cauchy distributed*, or to have the *Cauchy distribution,* if the density function of X is

$$f(x) = \frac{a}{\pi(x^2 + a^2)} \qquad a > 0, -\infty < x < \infty \tag{29}$$

This density function is symmetrical about $x = 0$ so that its median is zero. However, the mean, variance, and higher moments do not exist. Similarly, the moment generating function does not exist. However, the characteristic function does exist and is given by

$$\phi(\omega) = e^{-a\omega} \tag{30}$$

The Gamma Distribution

A random variable X is said to have the *gamma distribution*, or to be *gamma distributed*, if the density function is

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)} & x > 0 \\ 0 & x \le 0 \end{cases} \qquad (\alpha, \beta > 0) \tag{31}$$

where $\Gamma(\alpha)$ is the *gamma function* (see Appendix A). The mean and variance are given by

$$\mu = \alpha\beta, \qquad \sigma^2 = \alpha\beta^2 \tag{32}$$

The moment generating function and characteristic function are given, respectively, by

$$M(t) = (1 - \beta t)^{-\alpha}, \qquad \phi(\omega) = (1 - \beta i\omega)^{-\alpha} \tag{33}$$

The Beta Distribution

A random variable is said to have the *beta distribution*, or to be *beta distributed*, if the density function is

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1}(1 - x)^{\beta-1}}{B(\alpha, \beta)} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \qquad (\alpha, \beta > 0) \tag{34}$$

where $B(\alpha, \beta)$ is the *beta function* (see Appendix A). In view of the relation (9), Appendix A, between the beta and gamma functions, the beta distribution can also be defined by the density function

$$f(x) = \begin{cases} \dfrac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1 - x)^{\beta-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{35}$$

where α, β are positive. The mean and variance are

$$\mu = \frac{\alpha}{\alpha + \beta}, \qquad \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \tag{36}$$

For $\alpha > 1$, $\beta > 1$ there is a unique mode at the value

$$x_{\text{mode}} = \frac{\alpha - 1}{\alpha + \beta - 2} \tag{37}$$

The Chi-Square Distribution

Let $X_1, X_2, \ldots, X_v$ be v independent normally distributed random variables with mean zero and variance 1. Consider the random variable

$$\chi^2 = X_1^2 + X_2^2 + \cdots + X_v^2 \tag{38}$$

where χ^2 is called *chi square.* Then we can show that for $x \geq 0$,

$$P(\chi^2 \leq x) = \frac{1}{2^{v/2}\Gamma(v/2)} \int_0^x u^{(v/2)-1} e^{-u/2}\, du \tag{39}$$

and $P(\chi^2 \leq x) = 0$ for $x < 0$.

The distribution defined by (39) is called the *chi-square distribution*, and v is called the *number of degrees of freedom.* The distribution defined by (39) has corresponding density function given by

$$f(x) = \begin{cases} \dfrac{1}{2^{v/2}\Gamma(v/2)} x^{(v/2)-1} e^{-x/2} & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{40}$$

It is seen that the chi-square distribution is a special case of the gamma distribution with $\alpha = v/2$ $\beta = 2$. Therefore,

$$\mu = v, \qquad \sigma^2 = 2v, \qquad M(t) = (1 - 2t)^{-v/2}, \qquad \phi(\omega) = (1 - 2i\omega)^{-v/2} \tag{41}$$

For large $v (v \geq 30)$, we can show that $\sqrt{2\chi^2} - \sqrt{2v - 1}$ is very nearly normally distributed with mean 0 and variance 1.

Three theorems that will be useful in later work are as follows:

Theorem 4-3 Let $X_1, X_2, \ldots, X_v$ be independent normally distributed random variables with mean 0 and variance 1. Then $\chi^2 = X_1^2 + X_2^2 + \cdots + X_v^2$ is chi-square distributed with v degrees of freedom.

Theorem 4-4 Let $U_1, U_2, \ldots, U_k$ be independent random variables that are chi-square distributed with $v_1, v_2, \ldots, v_k$ degrees of freedom, respectively. Then their sum $W = U_1 + U_2 + \cdots + U_k$ is chi-square distributed with $v_1 + v_2 + \cdots + v_k$ degrees of freedom.

Theorem 4-5 Let V_1 and V_2 be independent random variables. Suppose that V_1 is chi-square distributed with v_1 degrees of freedom while $V = V_1 + V_2$ is chi-square distributed with v degrees of freedom, where $v > v_1$. Then V_2 is chi-square distributed with $v - v_1$ degrees of freedom.

In connection with the chi-square distribution, the t distribution (below), the F distribution (page 116), and others, it is common in statistical work to use the *same symbol* for both the random variable and a value of that random variable. Therefore, percentile values of the chi-square distribution for v degrees of freedom are denoted by $\chi^2_{p,v}$, or briefly χ^2_p if v is understood, and not by $x_{p,v}$ or x_p. (See Appendix E.) This is an ambiguous notation, and the reader should use care with it, especially when changing variables in density functions.

Student's *t* Distribution

If a random variable has the density function

$$f(t) = \frac{\Gamma\left(\dfrac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\dfrac{v}{2}\right)} \left(1 + \frac{t^2}{v}\right)^{-(v+1)/2} \qquad -\infty < t < \infty \tag{42}$$

it is said to have *Student's t distribution,* briefly the *t distribution*, with v degrees of freedom. If v is large ($v \geq 30$), the graph of $f(t)$ closely approximates the standard normal curve as indicated in Fig. 4-2. Percentile

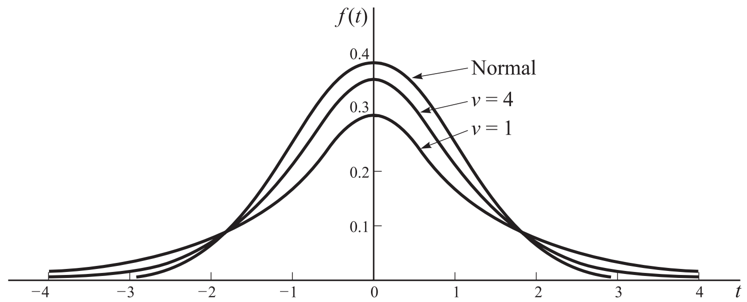


Fig. 4-2

values of the t distribution for v degrees of freedom are denoted by $t_{p,v}$ or briefly t_p if v is understood. For a table giving such values, see Appendix D. Since the t distribution is symmetrical, $t_{1-p} = -t_p$; for example, $t_{0.5} = -t_{0.95}$.

For the t distribution we have

$$\mu = 0 \qquad \text{and} \qquad \sigma^2 = \frac{v}{v-2} \qquad (v > 2). \tag{43}$$

The following theorem is important in later work.

Theorem 4-6 Let Y and Z be independent random variables, where Y is normally distributed with mean 0 and variance 1 while Z is chi-square distributed with v degrees of freedom. Then the random variable

$$T = \frac{Y}{\sqrt{Z/v}} \tag{44}$$

has the t distribution with v degrees of freedom.

The *F* Distribution

A random variable is said to have the *F distribution* (named after R. A. Fisher) *with* v_1 *and* v_2 *degrees of freedom* if its density function is given by

$$f(u) = \begin{cases} \dfrac{\Gamma\left(\dfrac{v_1 + v_2}{2}\right)}{\Gamma\left(\dfrac{v_1}{2}\right)\Gamma\left(\dfrac{v_2}{2}\right)} v_1^{v_1/2} v_2^{v_2/2} u^{(v_1/2)-1}(v_2 + v_1 u)^{-(v_1+v_2)/2} & u > 0 \\ 0 & u \le 0 \end{cases} \tag{45}$$

Percentile values of the F distribution for v_1, v_2 degrees of freedom are denoted by F_{p,v_1,v_2}, or briefly F_p if v_1, v_2 are understood. For a table giving such values in the case where $p = 0.95$ and $p = 0.99$, see Appendix F.

The mean and variance are given, respectively, by

$$\mu = \frac{v_2}{v_2 - 2} \ (v_2 > 2) \qquad \text{and} \qquad \sigma^2 = \frac{2v_2^2(v_1 + v_2 - 2)}{v_1(v_2-4)(v_2-2)^2} \qquad (v_2 > 4) \tag{46}$$

The distribution has a unique mode at the value

$$u_{\text{mode}} = \left(\frac{v_1 - 2}{v_1}\right)\left(\frac{v_2}{v_2 + 2}\right) \qquad (v_1 > 2) \tag{47}$$

The following theorems are important in later work.

Theorem 4-7 Let V_1 and V_2 be independent random variables that are chi-square distributed with v_1 and v_2 degrees of freedom, respectively. Then the random variable

$$V = \frac{V_1/v_1}{V_2/v_2} \tag{48}$$

has the F distribution with v_1 and v_2 degrees of freedom.

Theorem 4-8

$$F_{1-p,v_2,v_1} = \frac{1}{F_{p,v_1,v_2}}$$

Relationships Among Chi-Square, *t*, and *F* Distributions

Theorem 4-9 $$F_{1-p,1,v} = t^2_{1-(p/2),v}$$

Theorem 4-10 $$F_{p,v,\infty} = \frac{\chi^2_{p,v}}{v}$$

The Bivariate Normal Distribution

A generalization of the normal distribution to two continuous random variables X and Y is given by the joint density function

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\Big/2(1-\rho^2)\right\} \tag{49}$$

where $-\infty < x < \infty$, $-\infty < y < \infty$; μ_1, μ_2 are the means of X and Y; σ_1, σ_2 are the standard deviations of X and Y; and ρ is the correlation coefficient between X and Y. We often refer to (49) as the *bivariate normal distribution.*

For any joint distribution the condition $\rho = 0$ is necessary for independence of the random variables (see Theorem 3-15). In the case of (49) this condition is also sufficient (see Problem 4.51).

Miscellaneous Distributions

In the distributions listed below, the constants α, β, a, b, . . . are taken as positive unless otherwise stated. The characteristic function $\phi(\omega)$ is obtained from the moment generating function, where given, by letting $t = i\omega$.

1. GEOMETRIC DISTRIBUTION.

$$f(x) = P(X = x) = pq^{x-1} \qquad x = 1, 2, \ldots$$

$$\mu = \frac{1}{p} \qquad \sigma^2 = \frac{q}{p^2} \qquad M(t) = \frac{pe^t}{1-qe^t}$$

The random variable X represents the number of Bernoulli trials up to and including that in which the first success occurs. Here p is the probability of success in a single trial.

2. PASCAL'S OR NEGATIVE BINOMIAL DISTRIBUTION.

$$f(x) = P(X = x) = \binom{x-1}{r-1} p^r q^{x-r} \qquad x = r, r+1, \ldots$$

$$\mu = \frac{r}{P} \qquad \sigma^2 = \frac{rq}{p^2} \qquad M(t) = \left(\frac{pe^t}{1-qe^t}\right)^r$$

The random variable X represents the number of Bernoulli trials up to and including that in which the rth success occurs. The special case $r = 1$ gives the geometric distribution.

3. EXPONENTIAL DISTRIBUTION.

$$f(x) = \begin{cases} \alpha e^{-\alpha x} & x > 0 \\ 0 & x \le 0 \end{cases}$$

$$\mu = \frac{1}{\alpha} \qquad \sigma^2 = \frac{1}{\alpha^2} \qquad M(t) = \frac{\alpha}{\alpha - t}$$

4. WEIBULL DISTRIBUTION.

$$f(x) = \begin{cases} abx^{b-1}e^{-ax^b} & x > 0 \\ 0 & x \le 0 \end{cases}$$

$$\mu = a^{-1/b}\Gamma\left(1 + \frac{1}{b}\right) \qquad \sigma^2 = a^{-2/b}\left[\Gamma\left(1 + \frac{2}{b}\right) - \Gamma^2\left(1 + \frac{1}{b}\right)\right]$$

5. MAXWELL DISTRIBUTION.

$$f(x) = \begin{cases} \sqrt{2/\pi}\,\alpha^{3/2}x^2e^{-\alpha x^2/2} & x > 0 \\ 0 & x \le 0 \end{cases}$$

$$\mu = 2\sqrt{\frac{2}{\pi\alpha}} \qquad \sigma^2 = \left(3 - \frac{8}{\pi}\right)\alpha^{-1}$$

SOLVED PROBLEMS

The binomial distribution

4.1. Find the probability that in tossing a fair coin three times, there will appear (a) 3 heads, (b) 2 tails and 1 head, (c) at least 1 head, (d) not more than 1 tail.

Method 1

Let H denote heads and T denote tails, and suppose that we designate HTH, for example, to mean head on first toss, tail on second toss, and then head on third toss.

Since 2 possibilities (head or tail) can occur on each toss, there are a total of $(2)(2)(2) = 8$ possible outcomes, i.e., sample points, in the sample space. These are

$$HHH, HHT, HTH, HTT, TTH, THH, THT, TTT$$

For a fair coin these are assigned equal probabilities of $1/8$ each. Therefore,

(a) $P(\text{3 heads}) = P(HHH) = \dfrac{1}{8}$

(b) $P(\text{2 tails and 1 head}) = P(HTT \cup TTH \cup THT)$

$$= P(HTT) + P(TTH) + P(THT) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$

(c) $P(\text{at least 1 head})$

$$\begin{aligned} &= P(1, 2, \text{or } 3 \text{ heads}) \\ &= P(1\,\text{head}) + P(2\,\text{heads}) + P(3\text{ heads}) \\ &= P(HTT \cup THT \cup TTH) + P(HHT \cup HTH \cup THH) + P(HHH) \\ &= P(HTT) + P(THT) + P(TTH) + P(HHT) + P(HTH) + P(THH) + P(HHH) = \frac{7}{8} \end{aligned}$$

Alternatively,

$$P\,(\text{at least 1 head}) = 1 - P(\text{no head}) = 1 - P(TTT) = 1 - \frac{1}{8} = \frac{7}{8}$$

(d)
$$\begin{aligned} P(\text{not more than 1 tail}) &= P(\text{0 tails or 1 tail}) \\ &= P(\text{0 tails}) + P(\text{1 tail}) \\ &= P(HHH) + P(HHT \cup HTH \cup THH) \\ &= P(HHH) + P(HHT) + P(HTH) + P(THH) \\ &= \frac{4}{8} = \frac{1}{2} \end{aligned}$$

Method 2 (using formula)

(a) $P(3 \text{ heads}) = \binom{3}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^0 = \frac{1}{8}$

(b) $P(2 \text{ tails and 1 head}) = \binom{3}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^1 = \frac{3}{8}$

(c)
$$\begin{aligned} P(\text{at least 1 head}) &= P(1, 2, \text{ or } 3 \text{ heads}) \\ &= P(1 \text{ head}) + P(2 \text{ heads}) + P(3 \text{ heads}) \\ &= \binom{3}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^2 + \binom{3}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^1 + \binom{3}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^0 = \frac{7}{8} \end{aligned}$$

Alternatively,

$$\begin{aligned} P(\text{at least 1 head}) &= 1 - P(\text{no head}) \\ &= 1 - \binom{3}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^3 = \frac{7}{8} \end{aligned}$$

(d)
$$\begin{aligned} P(\text{not more than 1 tail}) &= P(0 \text{ tails or } 1 \text{ tail}) \\ &= P(0 \text{ tails}) + P(1 \text{ tail}) \\ &= \binom{3}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^0 + \binom{3}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right) = \frac{1}{2} \end{aligned}$$

It should be mentioned that the notation of random variables can also be used. For example, if we let X be the random variable denoting the number of heads in 3 tosses, (c) can be written

$$P(\text{at least 1 head}) = P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) = \frac{7}{8}$$

We shall use both approaches interchangeably.

4.2. Find the probability that in five tosses of a fair die, a 3 will appear (a) twice, (b) at most once, (c) at least two times.

Let the random variable X be the number of times a 3 appears in five tosses of a fair die. We have

$$\text{Probability of 3 in a single toss} = p = \frac{1}{6}$$

$$\text{Probability of no 3 in a single toss} = q = 1 - p = \frac{5}{6}$$

(a) $P(3 \text{ occurs twice}) = P(X = 2) = \binom{5}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 = \frac{625}{3888}$

(b)
$$\begin{aligned} P(3 \text{ occurs at most once}) &= P(X \leq 1) = P(X = 0) + P(X = 1) \\ &= \binom{5}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^5 + \binom{5}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^4 \\ &= \frac{3125}{7776} + \frac{3125}{7776} = \frac{3125}{3888} \end{aligned}$$

(c) $P(3 \text{ occurs at least 2 times})$

$$\begin{aligned} &= P(X \geq 2) \\ &= P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) \\ &= \binom{5}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 + \binom{5}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 + \binom{5}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^1 + \binom{5}{5}\left(\frac{1}{6}\right)^5\left(\frac{5}{6}\right)^0 \\ &= \frac{625}{3888} + \frac{125}{3888} + \frac{25}{7776} + \frac{1}{7776} = \frac{763}{3888} \end{aligned}$$

4.3. Find the probability that in a family of 4 children there will be (a) at least 1 boy, (b) at least 1 boy and at least 1 girl. Assume that the probability of a male birth is 1/2.

(a) $P(1\text{ boy}) = \binom{4}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 = \frac{1}{4}, \qquad P(2\text{ boys}) = \binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{3}{8}$

$P(3\text{ boys}) = \binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1 = \frac{1}{4}, \qquad P(4\text{ boys}) = \binom{4}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0 = \frac{1}{16}$

Then

$$P(\text{at least 1 boy}) = P(1\text{ boy}) + P(2\text{ boys}) + P(3\text{ boys}) + P(4\text{ boys})$$
$$= \frac{1}{4} + \frac{3}{8} + \frac{1}{4} + \frac{1}{16} = \frac{15}{16}$$

Another method

$$P(\text{at least 1 boy}) = 1 - P(\text{no boy}) = 1 - \left(\frac{1}{2}\right)^4 = 1 - \frac{1}{16} = \frac{15}{16}$$

(b) $P(\text{at least 1 boy and at least 1 girl}) = 1 - P(\text{no boy}) - P(\text{no girl})$

$$= 1 - \frac{1}{16} - \frac{1}{16} = \frac{7}{8}$$

We could also have solved this problem by letting X be a random variable denoting the number of boys in families with 4 children. Then, for example, (a) becomes

$$P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = \frac{15}{16}$$

4.4. Out of 2000 families with 4 children each, how many would you expect to have (a) at least 1 boy, (b) 2 boys, (c) 1 or 2 girls, (d) no girls?

Referring to Problem 4.3, we see that

(a) Expected number of families with at least 1 boy $= 2000\left(\frac{15}{16}\right) = 1875$

(b) Expected number of families with 2 boys $= 2000 \cdot P(2\text{ boys}) = 2000\left(\frac{3}{8}\right) = 750$

(c) $P(1\text{ or }2\text{ girls}) = P(1\text{ girl}) + P(2\text{ girls})$

$$= P(1\text{ boy}) + P(2\text{ boys}) = \frac{1}{4} + \frac{3}{8} = \frac{5}{8}$$

Expected number of families with 1 or 2 girls $= (2000)\left(\frac{5}{8}\right) = 1250$

(d) Expected number of families with no girls $= (2000)\left(\frac{1}{16}\right) = 125$

4.5. If 20% of the bolts produced by a machine are defective, determine the probability that out of 4 bolts chosen at random, (a) 1, (b) 0, (c) less than 2, bolts will be defective.

The probability of a defective bolt is $p = 0.2$, of a nondefective bolt is $q = 1 - p = 0.8$. Let the random variable X be the number of defective bolts. Then

(a) $P(X = 1) = \binom{4}{1}(0.2)^1(0.8)^3 = 0.4096$

(b) $P(X = 0) = \binom{4}{0}(0.2)^0(0.8)^4 = 0.4096$

(c) $P(X < 2) = P(X = 0) + P(X = 1)$
$= 0.4096 + 0.4096 = 0.8192$

4.6. Find the probability of getting a total of 7 at least once in three tosses of a pair of fair dice.

In a single toss of a pair of fair dice the probability of a 7 is $p = 1/6$ (see Problem 2.1, page 44), so that the probability of no 7 in a single toss is $q = 1 - p = 5/6$. Then

$$P(\text{no 7 in three tosses}) = \binom{3}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3 = \frac{125}{216}$$

and
$$P(\text{at least one 7 in three tosses}) = 1 - \frac{125}{216} = \frac{91}{216}$$

4.7. Find the moment generating function of a random variable X that is binomially distributed.

Method 1

If X is binomially distributed,

$$f(x) = P(X = x) = \binom{n}{x}p^x q^{n-x}$$

Then the moment generating function is given by

$$\begin{aligned} M(t) = E(e^{tx}) &= \sum e^{tx} f(x) \\ &= \sum_{x=0}^{n} e^{tx}\binom{n}{x} p^x q^{n-x} \\ &= \sum_{x=0}^{n}\binom{n}{x}(pe^t)^x q^{n-x} \\ &= (q + pe^t)^n \end{aligned}$$

Method 2

For a sequence of n Bernoulli trials, define

$$X_j = \begin{cases} 0 & \text{if failure in } j\text{th trial} \\ 1 & \text{if success in } j\text{th trial} \end{cases} \qquad (j = 1, 2, \dots, n)$$

Then the X_j are independent and $X = X_1 + X_2 + \cdots + X_n$. For the moment generating function of X_j, we have

$$M_j(t) = e^{t0}q + e^{t1}p = q + pe^t \qquad (j = 1, 2, \dots, n)$$

Then by Theorem 3-9, page 80,

$$M(t) = M_1(t)M_2(t)\dots M_n(t) = (q + pe^t)^n$$

4.8. Prove that the mean and variance of a binomially distributed random variable are, respectively, $\mu = np$ and $\sigma^2 = npq$.

Proceeding as in Method 2 of Problem 4.7, we have for $j = 1, 2, \dots, n$,

$$E(X_j) = 0q + 1p = p$$

$$\begin{aligned} \operatorname{Var}(X_j) &= E[(X_j - p)^2] = (0 - p)^2 q + (1 - p)^2 p \\ &= p^2 q + q^2 p = pq(p + q) = pq \end{aligned}$$

Then
$$\mu = E(X) = E(X_1) + E(X_2) + \cdots + E(X_n) = np$$

$$\sigma^2 = \operatorname{Var}(X) = \operatorname{Var}(X_1) + \operatorname{Var}(X_2) + \cdots + \operatorname{Var}(X_n) = npq$$

where we have used Theorem 3-7 for σ^2.

The above results can also be obtained (but with more difficulty) by differentiating the moment generating function (see Problem 3.38) or directly from the probability function.

4.9. If the probability of a defective bolt is 0.1, find (a) the mean, (b) the standard deviation, for the number of defective bolts in a total of 400 bolts.

(a) Mean $\mu = np = (400)\,(0.1) = 40$, i.e., we can *expect* 40 bolts to be defective.

(b) Variance $\sigma^2 = npq = (400)(0.1)(0.9) = 36$. Hence, the standard deviation $\sigma = \sqrt{36} = 6$.

The law of large numbers for Bernoulli trials

4.10. Prove Theorem 4-1, the (weak) law of large numbers for Bernoulli trials.

By Chebyshev's inequality, page 83, if X is any random variable with finite mean μ and variance σ^2, then

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \tag{1}$$

In particular, if X is binomially or Bernoulli distributed, then $\mu = np$, $\sigma = \sqrt{npq}$ and (1) becomes

$$P(|X - np| \geq k\sqrt{npq}) \leq \frac{1}{k^2} \tag{2}$$

or

$$P\left(\left|\frac{X}{n} - p\right| \geq k\sqrt{\frac{pq}{n}}\right) \leq \frac{1}{k^2} \tag{3}$$

If we let $\epsilon = k\sqrt{\frac{pq}{n}}$, (3) becomes

$$P\left(\left|\frac{X}{n} - p\right| \geq \epsilon\right) \leq \frac{pq}{n\epsilon^2}$$

and taking the limit as $n \to \infty$ we have, as required,

$$\lim_{n\to\infty} P\left(\left|\frac{X}{n} - p\right| \geq \epsilon\right) = 0$$

The result also follows directly from Theorem 3-19, page 83, with $S_n = X$, $\mu = np$, $\sigma = \sqrt{npq}$.

4.11. Give an interpretation of the (weak) law of large numbers for the appearances of a 3 in successive tosses of a fair die.

The law of large numbers states in this case that the probability of the proportion of 3s in n tosses differing from 1/6 by more than any value $\epsilon > 0$ approaches zero as $n \to \infty$.

The normal distribution

4.12. Find the area under the standard normal curve shown in Fig. 4-3 (a) between $z = 0$ and $z = 1.2$, (b) between $z = -0.68$ and $z = 0$, (c) between $z = -0.46$ and $z = 2.21$, (d) between $z = 0.81$ and $z = 1.94$, (e) to the right of $z = -1.28$.

(a) Using the table in Appendix C, proceed down the column marked z until entry 1.2 is reached. Then proceed right to column marked 0. The result, 0.3849, is the required area and represents the probability that Z is between 0 and 1.2 (Fig. 4-3). Therefore,

$$P(0 \leq Z \leq 1.2) = \frac{1}{\sqrt{2\pi}}\int_0^{1.2} e^{-u^2/2}\,du = 0.3849$$

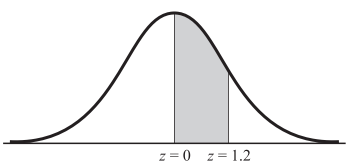


Fig. 4-3

(b) Required area = area between $z = 0$ and $z = +0.68$ (by symmetry). Therefore, proceed downward under column marked z until entry 0.6 is reached. Then proceed right to column marked 8.

The result, 0.2517, is the required area and represents the probability that Z is between -0.68 and 0 (Fig. 4-4). Therefore,

$$P(-0.68 \leq Z \leq 0) = \frac{1}{\sqrt{2\pi}}\int_{-0.68}^{0} e^{-u^2/2}\,du$$
$$= \frac{1}{\sqrt{2\pi}}\int_{0}^{0.68} e^{-u^2/2}\,du = 0.2517$$

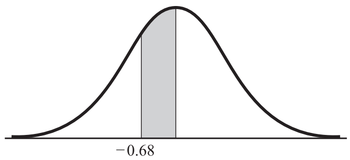


Fig. 4-4

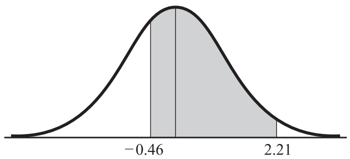


Fig. 4-5

(c) Required area = (area between $z = -0.46$ and $z = 0$)
+ (area between $z = 0$ and $z = 2.21$)
= (area between $z = 0$ and $z = 0.46$)
+ (area between $z = 0$ and $z = 2.21$)
= 0.1772 + 0.4864 = 0.6636

The area, 0.6636, represents the probability that Z is between -0.46 and 2.21 (Fig. 4-5). Therefore,

$$P(-0.46 \leq Z \leq 2.21) = \frac{1}{\sqrt{2\pi}}\int_{-0.46}^{2.21} e^{-u^2/2}\,du$$
$$= \frac{1}{\sqrt{2\pi}}\int_{-0.46}^{0} e^{-u^2/2}\,du + \frac{1}{\sqrt{2\pi}}\int_{0}^{2.21} e^{-u^2/2}\,du$$
$$= \frac{1}{\sqrt{2\pi}}\int_{0}^{0.46} e^{-u^2/2}\,du + \frac{1}{\sqrt{2\pi}}\int_{0}^{2.21} e^{-u^2/2}\,du = 0.1772 + 0.4864$$
$$= 0.6636$$

(d) Required area (Fig. 4-6) = (area between $z = 0$ and $z = 1.94$)
− (area between $z = 0$ and $z = 0.81$)
= 0.4738 − 0.2910 = 0.1828

This is the same as $P(0.81 \leq Z \leq 1.94)$.

(e) Required area (Fig. 4-7) = (area between $z = -1.28$ and $z = 0$)
+ (area to right of $z = 0$)
= 0.3997 + 0.5 = 0.8997

This is the same as $P(Z \geq -1.28)$.

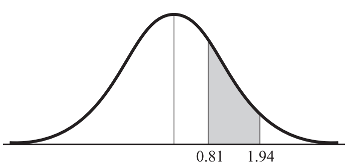


Fig. 4-6

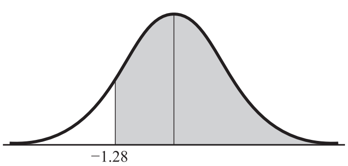


Fig. 4-7

4.13. If "area" refers to that under the standard normal curve, find the value or values of z such that (a) area between 0 and z is 0.3770, (b) area to left of z is 0.8621, (c) area between -1.5 and z is 0.0217.

(a) In the table in Appendix C the entry 0.3770 is located to the right of the row marked 1.1 and under the column marked 6. Then the required $z = 1.16$.

By symmetry, $z = -1.16$ is another value of z. Therefore, $z = \pm 1.16$ (Fig. 4-8). The problem is equivalent to solving for z the equation

$$\frac{1}{\sqrt{2\pi}}\int_0^z e^{-u^2/2}\,du = 0.3770$$

(b) Since the area is greater than 0.5, z must be positive.

Area between 0 and z is $0.8621 - 0.5 = 0.3621$, from which $z = 1.09$ (Fig. 4-9).

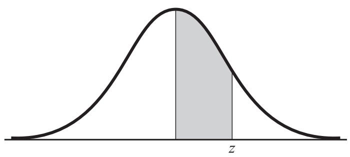


Fig. 4-8

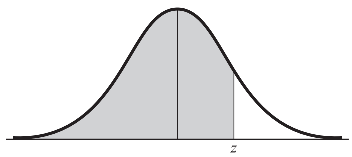


Fig. 4-9

(c) If z were positive, the area would be greater than the area between -1.5 and 0, which is 0.4332; hence z must be negative.

Case 1 z is negative but to the right of -1.5 (Fig. 4-10).

$$\begin{aligned}\text{Area between } -1.5 \text{ and } z &= (\text{area between } -1.5 \text{ and } 0) \\ &\quad - (\text{area between } 0 \text{ and } z) \\ 0.0217 &= 0.4332 - (\text{area between } 0 \text{ and } z)\end{aligned}$$

Then the area between 0 and z is $0.4332 - 0.0217 = 0.4115$ from which $z = -1.35$.

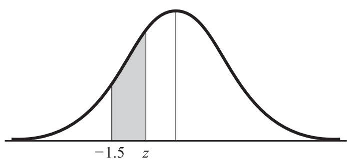


Fig. 4.10

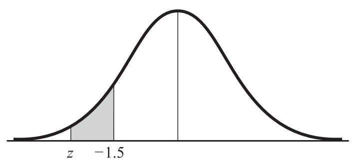


Fig. 4.11

Case 2 z is negative but to the left of -1.5 (Fig. 4-11).

$$\begin{aligned}\text{Area between } z \text{ and } -1.5 &= (\text{area between } z \text{ and } 0) \\ &\quad - (\text{area between } -1.5 \text{ and } 0) \\ 0.0217 &= (\text{area between } 0 \text{ and } z) - 0.4332\end{aligned}$$

Then the area between 0 and z is $0.0217 + 0.4332 = 0.4549$ and $z = -1.694$ by using linear interpolation; or, with slightly less precision, $z = -1.69$.

4.14. The mean weight of 500 male students at a certain college is 151 lb and the standard deviation is 15 lb. Assuming that the weights are normally distributed, find how many students weigh (a) between 120 and 155 lb, (b) more than 185 lb.

(a) Weights recorded as being between 120 and 155 lb can actually have any value from 119.5 to 155.5 lb, assuming they are recorded to the nearest pound (Fig. 4-12).

$$\begin{aligned}119.5 \text{ lb in standard units} &= (119.5 - 151)/15 \\ &= -2.10 \\ 155.5 \text{ lb in standard units} &= (155.5 - 151)/15 \\ &= 0.30\end{aligned}$$

$$\begin{aligned}\text{Required proportion of students} &= (\text{area between } z = -2.10 \text{ and } z = 0.30)\\ &= (\text{area between } z = -2.10 \text{ and } z = 0)\\ &\quad + (\text{area between } z = 0 \text{ and } z = 0.30)\\ &= 0.4821 + 0.1179 = 0.6000\end{aligned}$$

Then the number of students weighing between 120 and 155 lb is 500(0.6000) = 300

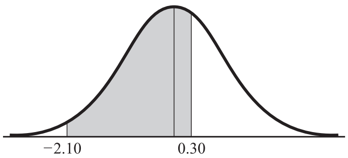


Fig. 4-12

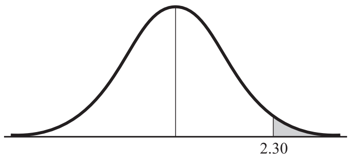


Fig. 4-13

(b) Students weighing more than 185 lb must weigh at least 185.5 lb (Fig. 4-13).

$$185.5 \text{ lb in standard units} = (185.5 - 151)/15 = 2.30$$

$$\begin{aligned}&\text{Required proportion of students}\\ &= (\text{area to right of } z = 2.30)\\ &= (\text{area to right of } z = 0)\\ &\quad - (\text{area between } z = 0 \text{ and } z = 2.30)\\ &= 0.5 - 0.4893 = 0.0107\end{aligned}$$

Then the number of students weighing more than 185 lb is 500(0.0107) = 5.

If W denotes the weight of a student chosen at random, we can summarize the above results in terms of probability by writing

$$P(119.5 \le W \le 155.5) = 0.6000 \qquad P(W \ge 185.5) = 0.0107$$

4.15. The mean inside diameter of a sample of 200 washers produced by a machine is 0.502 inches and the standard deviation is 0.005 inches. The purpose for which these washers are intended allows a maximum tolerance in the diameter of 0.496 to 0.508 inches, otherwise the washers are considered defective. Determine the percentage of defective washers produced by the machine, assuming the diameters are normally distributed.

$$0.496 \text{ in standard units} = (0.496 - 0.502)/0.005 = -1.2$$

$$0.508 \text{ in standard units} = (0.508 - 0.502)/0.005 = 1.2$$

Proportion of nondefective washers

$$\begin{aligned}&= (\text{area under normal curve between } z = -1.2 \text{ and } z = 1.2)\\ &= (\text{twice the area between } z = 0 \text{ and } z = 1.2)\\ &= 2(0.3849) = 0.7698, \text{ or } 77\%\end{aligned}$$

Therefore, the percentage of defective washers is 100% − 77% = 23% (Fig. 4-14).

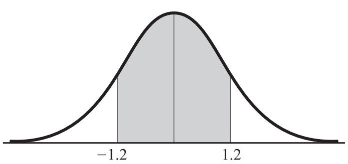


Fig. 4-14

Note that if we think of the interval 0.496 to 0.508 inches as actually representing diameters of from 0.4955 to 0.5085 inches, the above result is modified slightly. To two significant figures, however, the results are the same.

4.16. Find the moment generating function for the general normal distribution.

We have

$$M(t) = E(e^{tX}) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tx}e^{-(x-\mu)^2/2\sigma^2}\,dx$$

Letting $(x - \mu)/\sigma = v$ in the integral so that $x = \mu + \sigma v$, $dx = \sigma\, dv$, we have

$$M(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{ut+\sigma vt-(v^2/2)}\,dv = \frac{e^{\mu t+(\sigma^2t^2/2)}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(v-\sigma t)^2/2}\,dv$$

Now letting $v - \sigma t = w$, we find that

$$M(t) = e^{\mu t+(\sigma^2t^2/2)}\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-w^2/2}\,dw\right) = e^{ut+(\sigma^2t^2/2)}$$

Normal approximation to binomial distribution

4.17. Find the probability of getting between 3 and 6 heads inclusive in 10 tosses of a fair coin by using (a) the binomial distribution, (b) the normal approximation to the binomial distribution.

(a) Let X be the random variable giving the number of heads that will turn up in 10 tosses (Fig. 4-15). Then

$$P(X = 3) = \binom{10}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^7 = \frac{15}{128} \qquad P(X = 4) = \binom{10}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^6 = \frac{105}{512}$$

$$P(X = 5) = \binom{10}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^5 = \frac{63}{256} \qquad P(X = 6) = \binom{10}{6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^4 = \frac{105}{512}$$

Then the required probability is

$$P(3 \le X \le 6) = \frac{15}{128} + \frac{105}{512} + \frac{63}{256} + \frac{105}{512} = \frac{99}{128} = 0.7734$$

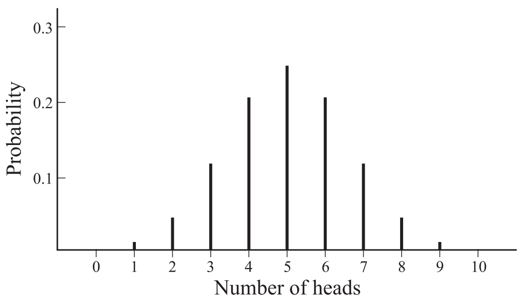


Fig. 4-15

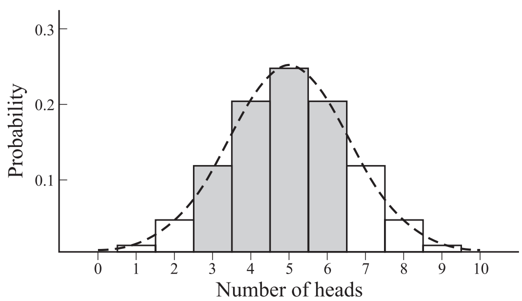


Fig. 4-16

(b) The probability distribution for the number of heads that will turn up in 10 tosses of the coin is shown graphically in Figures 4-15 and 4-16, where Fig. 4-16 treats the data as if they were continuous. The required probability is the sum of the areas of the shaded rectangles in Fig. 4-16 and can be approximated by the area under the corresponding normal curve, shown dashed. Treating the data as continuous, it follows that 3 to 6 heads can be considered as 2.5 to 6.5 heads. Also, the mean and variance for the binomial distribution are given by $\mu = np = 10\left(\frac{1}{2}\right) = 5$ and $\sigma = \sqrt{npq} = \sqrt{(10)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = 1.58$.

Now

$$2.5 \text{ in standard units} = \frac{2.5 - 5}{1.58} = -1.58$$

$$6.5 \text{ in standard units} = \frac{6.5 - 5}{1.58} = 0.95$$

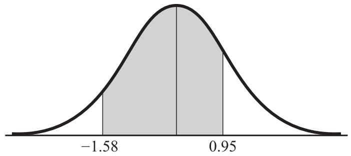


Fig. 4-17

$$\begin{aligned}\text{Required probability (Fig. 4-17)} &= (\text{area between } z = -1.58 \text{ and } z = 0.95)\\ &= (\text{area between } z = -1.58 \text{ and } z = 0)\\ &\quad + (\text{area between z} = 0 \text{ and } z = 0.95)\\ &= 0.4429 + 0.3289 = 0.7718\end{aligned}$$

which compares very well with the true value 0.7734 obtained in part (a). The accuracy is even better for larger values of n.

4.18. A fair coin is tossed 500 times. Find the probability that the number of heads will not differ from 250 by (a) more than 10, (b) more than 30.

$$\mu = np = (500)\left(\frac{1}{2}\right) = 250 \qquad \sigma = \sqrt{npq} = \sqrt{(500)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = 11.18$$

(a) We require the probability that the number of heads will lie between 240 and 260, or considering the data as continuous, between 239.5 and 260.5.

$$239.5 \text{ in standard units } = \frac{239.5 - 250}{11.18} = -0.94 \qquad 260.5 \text{ in standard units} = 0.94$$

$$\begin{aligned}\text{Required probability} &= (\text{area under normal curve between } z = -0.94 \text{ and } z = 0.94)\\ &= (\text{twice area between } z = 0 \text{ and } z = 0.94) = 2(0.3264) = 0.6528\end{aligned}$$

(b) We require the probability that the number of heads will lie between 220 and 280 or, considering the data as continuous, between 219.5 and 280.5.

$$219.5 \text{ in standard units } = \frac{219.5 - 250}{11.18} = -2.73 \qquad 280.5 \text{ in standard units} = 2.73$$

$$\begin{aligned}\text{Required probability} &= (\text{twice area under normal curve between } z = 0 \text{ and } z = 2.73)\\ &= 2(0.4968) = 0.9936\end{aligned}$$

It follows that we can be very confident that the number of heads will not differ from that expected (250) by more than 30. Therefore, if it turned out that the *actual* number of heads was 280, we would strongly believe that the coin was not fair, i.e., it was *loaded.*

4.19. A die is tossed 120 times. Find the probability that the face 4 will turn up (a) 18 times or less, (b) 14 times or less, assuming the die is fair.

The face 4 has probability $p = \frac{1}{6}$ of turning up and probability $q = \frac{5}{6}$ of not turning up.

(*a*) We want the probability of the number of 4s being between 0 and 18. This is given exactly by

$$\binom{120}{18}\left(\frac{1}{2}\right)^{18}\left(\frac{5}{6}\right)^{102} + \binom{120}{17}\left(\frac{1}{6}\right)^{17}\left(\frac{5}{6}\right)^{103} + \cdots + \binom{120}{0}\left(\frac{1}{6}\right)^{0}\left(\frac{5}{6}\right)^{120}$$

but since the labor involved in the computation is overwhelming, we use the normal approximation.

Considering the data as continuous, it follows that 0 to 18 4s can be treated as −0.5 to 18.5 4s. Also,

$$\mu = np = 120\left(\frac{1}{6}\right) = 20 \qquad \text{and} \qquad \sigma = \sqrt{npq} = \sqrt{(120)\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)} = 4.08$$

Then

$$-0.5 \text{ in standard units} = \frac{-0.5 - 20}{4.08} = -5.02. \qquad 18.5 \text{ in standard units} = -0.37$$

$$\begin{aligned} \text{Required probability} &= (\text{area under normal curve between } z = -5.02 \text{ and } z = -0.37) \\ &= (\text{area between } z = 0 \text{ and } z = -5.02) \\ &\quad - (\text{area between } z = 0 \text{ and } z = -0.37) \\ &= 0.5 - 0.1443 = 0.3557 \end{aligned}$$

(b) We proceed as in part (a), replacing 18 by 14. Then

$$-0.5 \text{ in standard units} = -5.02 \qquad 14.5 \text{ in standard units} = \frac{14.5 - 20}{4.08} = -1.35$$

$$\begin{aligned} \text{Required probability} &= (\text{area under normal curve between } z = -5.02 \text{ and } z = -1.35) \\ &= (\text{area between } z = 0 \text{ and } z = -5.02) \\ &\quad - (\text{area between } z = 0 \text{ and } z = -1.35) \\ &= 0.5 - 0.4115 = 0.0885 \end{aligned}$$

It follows that if we were to take repeated samples of 120 tosses of a die, a 4 should turn up 14 times or less in about one-tenth of these samples.

The Poisson distribution

4.20. Establish the validity of the Poisson approximation to the binomial distribution.

If X is binomially distributed, then

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \tag{1}$$

where $E(X) = np$. Let $\lambda = np$ so that $p = \lambda/n$. Then (1) becomes

$$\begin{aligned} P(X = x) &= \binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{n(n-1)(n-2)\cdots(n-x+1)}{x!n^x}\lambda^x\left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{x-1}{n}\right)}{x!}\lambda^x\left(1 - \frac{\lambda}{n}\right)^{n-x} \end{aligned}$$

Now as $n \to \infty$,

$$\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{x-1}{n}\right) \to 1$$

while

$$\left(1 - \frac{\lambda}{n}\right)^{n-x} = \left(1 - \frac{\lambda}{n}\right)^n\left(1 - \frac{\lambda}{n}\right)^{-x} \to (e^{-\lambda})(1) = e^{-\lambda}$$

using the well-known result from calculus that

$$\lim_{n\to\infty}\left(1 + \frac{u}{n}\right)^n = e^u$$

It follows that when $n \to \infty$ but λ stays fixed (i.e., $p \to 0$),

$$P(X = x) \to \frac{\lambda^x e^{-\lambda}}{x!} \tag{2}$$

which is the Poisson distribution.

Another method

The moment generating function for the binomial distribution is

(3) $$(q + pe^t)^n = (1 - p + pe^t)^n = [1 + p(e^t - 1)]^n$$

If $\lambda = np$ so that $p = \lambda/n$, this becomes

(4) $$\left[1 + \frac{\lambda(e^t - 1)}{n}\right]^n$$

As $n \to \infty$ this approaches

(5) $$e^{\lambda(e^t - 1)}$$

which is the moment generating function of the Poisson distribution. The required result then follows on using Theorem 3-10, page 77.

4.21. Verify that the limiting function (2) of Problem 4.20 is actually a probability function.

First, we see that $P(X = x) > 0$ for $x = 0, 1, \ldots,$ given that $\lambda > 0$. Second, we have

$$\sum_{x=0}^{\infty} P(X = x) = \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} \cdot e^{\lambda} = 1$$

and the verification is complete.

4.22. Ten percent of the tools produced in a certain manufacturing process turn out to be defective. Find the probability that in a sample of 10 tools chosen at random, exactly 2 will be defective, by using (a) the binomial distribution, (b) the Poisson approximation to the binomial distribution.

(a) The probability of a defective tool is $p = 0.1$. Let X denote the number of defective tools out of 10 chosen. Then, according to the binomial distribution,

$$P(X = 2) = \binom{10}{2}(0.1)^2(0.9)^8 = 0.1937 \text{ or } 0.19$$

(b) We have $\lambda = np = (10)(0.1) = 1$. Then, according to the Poisson distribution,

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad \text{or} \qquad P(X = 2) = \frac{(1)^2 e^{-1}}{2!} = 0.1839 \text{ or } 0.18$$

In general, the approximation is good if $p \le 0.1$ and $\lambda = np \le 5$.

4.23. If the probability that an individual will suffer a bad reaction from injection of a given serum is 0.001, determine the probability that out of 2000 individuals, (a) exactly 3, (b) more than 2, individuals will suffer a bad reaction.

Let X denote the number of individuals suffering a bad reaction. X is Bernoulli distributed, but since bad reactions are assumed to be rare events, we can suppose that X is Poisson distributed, i.e.,

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad \text{where } \lambda = np = (2000)(0.001) = 2$$

(a) $$P(X = 3) = \frac{2^3 e^{-2}}{3!} = 0.180$$

(b) $$\begin{aligned} P(X > 2) &= 1 - [P(X = 0) + P(X = 1) + P(X = 2)] \\ &= 1 - \left[\frac{2^0 e^{-2}}{0!} + \frac{2^1 e^{-2}}{1!} + \frac{2^2 e^{-2}}{2!}\right] \\ &= 1 - 5e^{-2} = 0.323 \end{aligned}$$

An exact evaluation of the probabilities using the binomial distribution would require much more labor.

The central limit theorem

4.24. Verify the central limit theorem for a random variable X that is binomially distributed, and thereby establish the validity of the normal approximation to the binomial distribution.

The standardized variable for X is $X^* = (X - np)/\sqrt{npq}$, and the moment generating function for X^* is

$$\begin{aligned}
E(e^{tX^*}) &= E(e^{t(X-np)/\sqrt{npq}}) \\
&= e^{-tnp/\sqrt{npq}}\, E(e^{tX/\sqrt{npq}}) \\
&= e^{-tnp/\sqrt{npq}} \sum_{x=0}^{n} e^{tx/\sqrt{npq}} \binom{n}{x} p^x q^{n-x} \\
&= e^{-tnp/\sqrt{npq}} \sum_{x=0}^{n} \binom{n}{x} (pe^{t/\sqrt{npq}})^x q^{n-x} \\
&= e^{-tnp/\sqrt{npq}} (q + pe^{t/\sqrt{npq}})^n \\
&= [e^{-tp/\sqrt{npq}} (q + pe^{t/\sqrt{npq}})]^n \\
&= (qe^{-tp/\sqrt{npq}} + pe^{tq/\sqrt{npq}})^n
\end{aligned}$$

Using the expansion

$$e^u = 1 + u + \frac{u^2}{2!} + \frac{u^3}{3!} + \cdots$$

we find

$$\begin{aligned}
qe^{-tp/\sqrt{npq}} + pe^{tq/\sqrt{npq}} &= q\left(1 - \frac{tp}{\sqrt{npq}} + \frac{t^2p^2}{2npq} + \cdots\right) \\
&\quad + p\left(1 + \frac{tq}{\sqrt{npq}} + \frac{t^2q^2}{2npq} + \cdots\right) \\
&= q + p + \frac{pq(p + q)t^2}{2npq} + \cdots \\
&= 1 + \frac{t^2}{2n} + \cdots
\end{aligned}$$

Therefore, $$E(e^{tX^*}) = \left(1 + \frac{t^2}{2n} + \cdots\right)^n$$

But as $n \to \infty$, the right-hand side approaches $e^{t^2/2}$, which is the moment generating function for the standard normal distribution. Therefore, the required result follows by Theorem 3-10, page 77.

4.25. Prove the central limit theorem (Theorem 4-2, page 112).

For $n = 1, 2, \ldots$, we have $S_n = X_1 + X_2 + \cdots + X_n$. Now $X_1, X_2, \ldots, X_n$ each have mean μ and variance σ^2. Thus

$$E(S_n) = E(X_1) + E(X_2) + \cdots + E(X_n) = n\mu$$

and, because the X_k are independent,

$$\mathrm{Var}(S_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) = n\sigma^2$$

It follows that the standardized random variable corresponding to S_n is

$$S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

The moment generating function for S_n^* is

$$\begin{aligned}
E(e^{tS_n^*}) &= E[e^{t(S_n - n\mu)/\sigma\sqrt{n}}] \\
&= E[e^{t(X_1-\mu)/\sigma\sqrt{n}} e^{t(X_2-\mu)/\sigma\sqrt{n}} \cdots e^{t(X_n-\mu)/\sigma\sqrt{n}}] \\
&= E[e^{t(X_1-\mu)/\sigma\sqrt{n}}] \cdot E[e^{t(X_2-\mu)/\sigma\sqrt{n}}] \cdots E[e^{t(X_n-\mu)/\sigma\sqrt{n}}] \\
&= \{E[e^{t(X_1-\mu)/\sigma\sqrt{n}}]\}^n
\end{aligned}$$

where, in the last two steps, we have respectively used the facts that the X_k are independent and are identically distributed. Now, by a Taylor series expansion,

$$E[e^{t(X_1-\mu)/\sigma\sqrt{n}}] = E\left[1 + \frac{t(X_1 - \mu)}{\sigma\sqrt{n}} + \frac{t^2(X_1 - \mu)^2}{2\sigma^2 n} + \cdots\right]$$

$$= E(1) + \frac{t}{\sigma\sqrt{n}}E(X_1 - \mu) + \frac{t^2}{2\sigma^2 n}E[(X_1 - \mu)^2] + \cdots$$

$$= 1 + \frac{t}{\sigma\sqrt{n}}(0) + \frac{t^2}{2\sigma^2 n}(\sigma^2) + \cdots = 1 + \frac{t^2}{2n} + \cdots$$

so that

$$E(e^{tS_n^*}) = \left(1 + \frac{t^2}{2n} + \cdots\right)^n$$

But the limit of this as $n \to \infty$ is $e^{t^2/2}$, which is the moment generating function of the standardized normal distribution. Hence, by Theorem 3-10, page 80, the required result follows.

Multinomial distribution

4.26. A box contains 5 red balls, 4 white balls, and 3 blue balls. A ball is selected at random from the box, its color is noted, and then the ball is replaced. Find the probability that out of 6 balls selected in this manner, 3 are red, 2 are white, and 1 is blue.

Method 1 (by formula)

$$P(\text{red at any drawing}) = \frac{5}{12} \qquad P(\text{white at any drawing}) = \frac{4}{12}$$

$$P(\text{blue at any drawing}) = \frac{3}{12}$$

Then

$$P(3 \text{ red}, 2 \text{ white}, 1 \text{ blue}) = \frac{6!}{3!2!1!}\left(\frac{5}{12}\right)^3\left(\frac{4}{12}\right)^2\left(\frac{3}{12}\right)^1 = \frac{625}{5184}$$

Method 2

The probability of choosing any red ball is $5/12$. Then the probability of choosing 3 red balls is $(5/12)^3$. Similarly, the probability of choosing 2 white balls is $(4/12)^2$, and of choosing 1 blue ball, $(3/12)^1$. Therefore, the probability of choosing 3 red, 2 white, and 1 blue in that order is

$$\left(\frac{5}{12}\right)^3\left(\frac{4}{12}\right)^2\left(\frac{3}{12}\right)^1$$

But the same selection can be achieved in various other orders, and the number of these different ways is

$$\frac{6!}{3!2!1!}$$

as shown in Chapter 1. Then the required probability is

$$\frac{6!}{3!2!1!}\left(\frac{5}{12}\right)^3\left(\frac{4}{12}\right)^2\left(\frac{3}{12}\right)^1$$

Method 3

The required probability is the term $p_r^3 p_w^2 p_b$ in the multinomial expansion of $(p_r + p_w + p_b)^6$ where $p_r = 5/12$, $p_w = 4/12$, $p_b = 3/12$. By actual expansion, the above result is obtained.

The hypergeometric distribution

4.27. A box contains 6 blue marbles and 4 red marbles. An experiment is performed in which a marble is chosen at random and its color observed, but the marble is not replaced. Find the probability that after 5 trials of the experiment, 3 blue marbles will have been chosen.

Method 1

The number of different ways of selecting 3 blue marbles out of 6 blue marbles is $\binom{6}{3}$. The number of different ways of selecting the remaining 2 marbles out of the 4 red marbles is $\binom{4}{2}$. Therefore, the number of different samples containing 3 blue marbles and 2 red marbles is $\binom{6}{3}\binom{4}{2}$.

Now the total number of different ways of selecting 5 marbles out of the 10 marbles (6 + 4) in the box is $\binom{10}{5}$. Therefore, the required probability is given by

$$\frac{\binom{6}{3}\binom{4}{2}}{\binom{10}{5}} = \frac{10}{21}$$

Method 2 (using formula)
We have $b = 6$, $r = 4$, $n = 5$, $x = 3$. Then by (*19*), page 113, the required probability is

$$P(X = 3) = \frac{\binom{6}{3}\binom{4}{2}}{\binom{10}{2}}$$

The uniform distribution

4.28. Show that the mean and variance of the uniform distribution (page 113) are given respectively by (a) $\mu = \frac{1}{2}(a + b)$, (b) $\sigma^2 = \frac{1}{12}(b - a)^2$.

(a) $$\mu = E(X) = \int_a^b \frac{x\,dx}{b - a} = \frac{x^2}{2(b - a)}\bigg|_a^b = \frac{b^2 - a^2}{2(b - a)} = \frac{a + b}{2}$$

(b) We have

$$E(X^2) = \int_a^b \frac{x^2 dx}{b - a} = \frac{x^3}{3(b - a)}\bigg|_a^b = \frac{b^3 - a^3}{3(b - a)} = \frac{b^2 + ab + a^2}{3}$$

Then the variance is given by

$$\sigma^2 = E[(X - \mu)^2] = E(X^2) - \mu^2$$
$$= \frac{b^2 + ab + a^2}{3} - \left(\frac{a + b}{2}\right)^2 = \frac{1}{12}(b - a)^2$$

The Cauchy distribution

4.29. Show that (a) the moment generating function for a Cauchy distributed random variable X does not exist but that (b) the characteristic function does exist.

(a) The moment generating function of X is

$$E(e^{tX}) = \frac{a}{\pi}\int_{-\infty}^{\infty} \frac{e^{tx}}{x^2 + a^2}\,dx$$

which does not exist if t is real. This can be seen by noting, for example, that if $x \geq 0$, $t > 0$,

$$e^{tx} = 1 + tx + \frac{t^2x^2}{2!} + \cdots > \frac{t^2x^2}{2}$$

so that

$$\frac{a}{\pi}\int_{-\infty}^{\infty} \frac{e^{tx}}{x^2 + a^2}\,dx \geq \frac{at^2}{2\pi}\int_0^{\infty} \frac{x^2}{x^2 + a^2}\,dx$$

and the integral on the right diverges.

(b) The characteristic function of X is

$$E(e^{i\omega X}) = \frac{a}{\pi}\int_{-\infty}^{\infty} \frac{e^{i\omega x}}{x^2 + a^2}\,dx$$
$$= \frac{a}{\pi}\int_{-\infty}^{\infty} \frac{\cos \omega x}{x^2 + a^2}\,dx + \frac{ai}{\pi}\int_{-\infty}^{\infty} \frac{\sin \omega x}{x^2 + a^2}\,dx$$
$$= \frac{2a}{\pi}\int_0^{\infty} \frac{\cos \omega x}{x^2 + a^2}\,dx$$

where we have used the fact that the integrands in the next to last line are even and odd functions, respectively. The last integral can be shown to exist and to equal $e^{-a\omega}$.

4.30. Let Θ be a uniformly distributed random variable in the interval $-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$. Prove that $X = a \tan \Theta$, $a > 0$, is Cauchy distributed in $-\infty < x < \infty$.

The density function of Θ is

$$f(\theta) = \frac{1}{\pi} \qquad -\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$$

Considering the transformation $x = a \tan \theta$, we have

$$\theta = \tan^{-1}\frac{x}{a} \qquad \text{and} \qquad \frac{d\theta}{dx} = \frac{a}{x^2 + a^2} > 0$$

Then by Theorem 2-3, page 42, the density function of X is given by

$$g(x) = f(\theta)\left|\frac{d\theta}{dx}\right| = \frac{1}{\pi}\frac{a}{x^2 + a^2}$$

which is the Cauchy distribution.

The gamma distribution

4.31. Show that the mean and variance of the gamma distribution are given by (a) $\mu = \alpha\beta$, (b) $\sigma^2 = \alpha\beta^2$.

(a) $$\mu = \int_0^\infty x\left[\frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}\right]dx = \int_0^\infty \frac{x^\alpha e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}\,dx$$

Letting $x/\beta = t$, we have

$$\mu = \frac{\beta^\alpha\beta}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty t^\alpha e^{-t}\,dt = \frac{\beta}{\Gamma(\alpha)}\Gamma(\alpha + 1) = \alpha\beta$$

(b) $$E(X^2) = \int_0^\infty x^2\left[\frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}\right]dx = \int_0^\infty \frac{x^{\alpha+1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}\,dx$$

Letting $x/\beta = t$, we have

$$\begin{aligned} E(X^2) &= \frac{\beta^{\alpha+1}\beta}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty t^{\alpha+1}e^{-t}\,dt \\ &= \frac{\beta^2}{\Gamma(\alpha)}\Gamma(\alpha + 2) = \beta^2(\alpha + 1)\alpha \end{aligned}$$

since $\Gamma(\alpha + 2) = (\alpha + 1)\Gamma(\alpha + 1) = (\alpha + 1)\alpha\Gamma(\alpha)$. Therefore,

$$\sigma^2 = E(X^2) - \mu^2 = \beta^2(\alpha + 1)\alpha - (\alpha\beta)^2 = \alpha\beta^2$$

The beta distribution

4.32. Find the mean of the beta distribution.

$$\begin{aligned} \mu = E(X) &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 x[x^{\alpha-1}(1 - x)^{\beta-1}]\,dx \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 x^\alpha(1 - x)^{\beta-1}\,dx \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\frac{\Gamma(\alpha + 1)\Gamma(\beta)}{\Gamma(\alpha + 1 + \beta)} \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha + \beta)\Gamma(\alpha + \beta)} = \frac{\alpha}{\alpha + \beta} \end{aligned}$$

4.33. Find the variance of the beta distribution.

The second moment about the origin is

$$E(X^2) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 x^2[x^{\alpha-1}(1-x)^{\beta-1}]\,dx$$
$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 x^{\alpha+1}(1-x)^{\beta-1}\,dx$$
$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+2+\beta)}$$
$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\frac{(\alpha+1)\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)}$$
$$= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

Then using Problem 4.32, the variance is

$$\sigma^2 = E(X^2) - [E(X)]^2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

The chi-square distribution

4.34. Show that the moment generating function of a random variable X, which is chi-square distributed with v degrees of freedom, is $M(t) = (1-2t)^{-v/2}$.

$$M(t) = E(e^{tX}) = \frac{1}{2^{v/2}\Gamma(v/2)}\int_0^\infty e^{tx}x^{(v-2)/2}e^{-x/2}\,dx$$
$$= \frac{1}{2^{v/2}\Gamma(v/2)}\int_0^\infty x^{(v-2)/2}e^{-(1-2t)x/2}\,dx$$

Letting $(1-2t)x/2 = u$ in the last integral, we find

$$M(t) = \frac{1}{2^{v/2}\Gamma(v/2)}\int_0^\infty \left(\frac{2u}{1-2t}\right)^{(v-2)/2} e^{-u}\,\frac{2\,du}{1-2t}$$
$$= \frac{(1-2t)^{-v/2}}{\Gamma(v/2)}\int_0^\infty u^{(v/2)-1}e^{-u}\,du = (1-2t)^{-v/2}$$

4.35. Let X_1 and X_2 be independent random variables that are chi-square distributed with v_1 and v_2 degrees of freedom, respectively, (a) Show that the moment generating function of $Z = X_1 + X_2$ is $(1-2t)^{-(v_1+v_2)/2}$, thereby (b) show that Z is chi-square distributed with $v_1 + v_2$ degrees of freedom.

(a) The moment generating function of $Z = X_1 + X_2$ is

$$M(t) = E[e^{t(X_1+X_2)}] = E(e^{tX_1})E(e^{tX_2}) = (1-2t)^{-v_1/2}(1-2t)^{-v_2/2} = (1-2t)^{-(v_1+v_2)/2}$$

using Problem 4.34.

(b) It is seen from Problem 4.34 that a distribution whose moment generating function is $(1-2t)^{-(v_1+v_2)/2}$ is the chi-square distribution with $v_1 + v_2$ degrees of freedom. This must be the distribution of Z, by Theorem 3-10, page 77.

By generalizing the above results, we obtain a proof of Theorem 4-4, page 115.

4.36. Let X be a normally distributed random variable having mean 0 and variance 1. Show that X^2 is chi-square distributed with 1 degree of freedom.

We want to find the distribution of $Y = X^2$ given a standard normal distribution for X. Since the correspondence between X and Y is not one-one, we cannot apply Theorem 2-3 as it stands but must proceed as follows.

For $y < 0$, it is clear that $P(Y \le y) = 0$. For $y \ge 0$, we have

$$P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le +\sqrt{y})$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{y}}^{+\sqrt{y}} e^{-x^2/2}\,dx = \frac{2}{\sqrt{2\pi}} \int_{0}^{+\sqrt{y}} e^{-x^2/2}\,dx$$

where the last step uses the fact that the standard normal density function is even. Making the change of variable $x = +\sqrt{t}$ in the final integral, we obtain

$$P(Y \le y) = \frac{1}{\sqrt{2\pi}} \int_0^y t^{-1/2} e^{-t/2}\,dt$$

But this is a chi-square distribution with 1 degree of freedom, as is seen by putting $v = 1$ in (39), page 115, and using the fact that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

4.37. Prove Theorem 4-3, page 115, for $v = 2$.

By Problem 4.36 we see that if X_1 and X_2 are normally distributed with mean 0 and variance 1, then X_1^2 and X_2^2 are chi square distributed with 1 degree of freedom each. Then, from Problem 4.35(b), we see that $Z = X_1^2 + X_2^2$ is chi square distributed with $1 + 1 = 2$ degrees of freedom if X_1 and X_2 are independent. The general result for all positive integers v follows in the same manner.

4.38. The graph of the chi-square distribution with 5 degrees of freedom is shown in Fig. 4-18. (See the remarks on notation on page 115.) Find the values χ_1^2, χ_2^2 for which

(a) the shaded area on the right = 0.05,

(b) the total shaded area = 0.05,

(c) the shaded area on the left = 0.10,

(d) the shaded area on the right = 0.01.

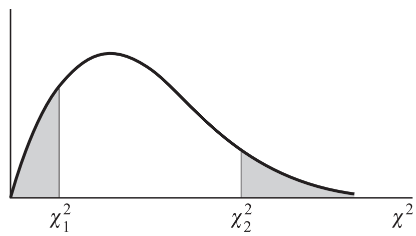


Fig. 4-18

(a) If the shaded area on the right is 0.05, then the area to the left of χ_2^2 is $(1 - 0.05) = 0.95$, and χ_2^2 represents the 95th percentile, $\chi_{0.95}^2$.

Referring to the table in Appendix E, proceed downward under column headed v until entry 5 is reached. Then proceed right to the column headed $\chi_{0.95}^2$. The result, 11.1, is the required value of χ^2.

(b) Since the distribution is not symmetric, there are many values for which the total shaded area = 0.05. For example, the right-hand shaded area could be 0.04 while the left-hand shaded area is 0.01. It is customary, however, unless otherwise specified, to choose the two areas equal. In this case, then, each area = 0.025.

If the shaded area on the right is 0.025, the area to the left of χ_2^2 is $1 - 0.025 = 0.975$ and χ_2^2 represents the 97.5th percentile, $\chi_{0.975}^2$, which from Appendix E is 12.8.

Similarly, if the shaded area on the left is 0.025, the area to the left of χ_1^2 is 0.025 and χ_1^2 represents the 2.5th percentile, $\chi_{0.025}^2$, which equals 0.831.

Therefore, the values are 0.831 and 12.8.

(c) If the shaded area on the left is 0.10, χ_1^2 represents the 10th percentile, $\chi_{0.10}^2$, which equals 1.61.

(d) If the shaded area on the right is 0.01, the area to the left of χ_2^2 is 0.99, and χ_2^2 represents the 99th percentile, $\chi_{0.99}^2$, which equals 15.1.

4.39. Find the values of χ^2 for which the area of the right-hand tail of the χ^2 distribution is 0.05, if the number of degrees of freedom v is equal to (a) 15, (b) 21, (c) 50.

Using the table in Appendix E, we find in the column headed $\chi^2_{0.95}$ the values: (a) 25.0 corresponding to $v = 15$; (b) 32.7 corresponding to $v = 21$; (c) 67.5 corresponding to $v = 50$.

4.40. Find the median value of χ^2 corresponding to (a) 9, (b) 28, (c) 40 degrees of freedom.

Using the table in Appendix E, we find in the column headed $\chi^2_{0.50}$ (since the median is the 50th percentile) the values: (a) 8.34 corresponding to $v = 9$; (b) 27.3 corresponding to $v = 28$; (c) 39.3 corresponding to $v = 40$.

It is of interest to note that the median values are very nearly equal to the number of degrees of freedom. In fact, for $v > 10$ the median values are equal to $v - 0.7$, as can be seen from the table.

4.41. Find $\chi^2_{0.95}$ for (a) $v = 50$, (b) $v = 100$ degrees of freedom.

For v greater than 30, we can use the fact that $(\sqrt{2\chi^2} - \sqrt{2v - 1})$ is very closely normally distributed with mean zero and variance one. Then if z_p is the $(100p)$th percentile of the standardized normal distribution, we can write, to a high degree of approximation,

$$\sqrt{2\chi^2_p} - \sqrt{2v - 1} = z_p \qquad \text{or} \qquad \sqrt{2\chi^2_p} = z_p + \sqrt{2v - 1}$$

from which

$$\chi^2_p = \tfrac{1}{2}(z_p + \sqrt{2v - 1})^2$$

(a) If $v = 50$, $\chi^2_{0.95} = \frac{1}{2}(z_{0.95} + \sqrt{2(50) - 1})^2 = \frac{1}{2}(1.64 + \sqrt{99})^2 = 69.2$, which agrees very well with the value 67.5 given in Appendix E.

(b) If $v = 100$, $\chi^2_{0.95} = \frac{1}{2}(z_{0.95} + \sqrt{2(100) - 1})^2 = \frac{1}{2}(1.64 + \sqrt{199})^2 = 124.0$ (actual value $= 124.3$).

Student's *t* distribution

4.42. Prove Theorem 4-6, page 116.

Since Y is normally distributed with mean 0 and variance 1, its density function is

$$\frac{1}{\sqrt{2\pi}} e^{-y^2/2} \tag{1}$$

Since Z is chi-square distributed with v degrees of freedom, its density function is

$$\frac{1}{2^{v/2}\Gamma(v/2)} z^{(v/2)-1} e^{-z/2} \qquad z > 0 \tag{2}$$

Because Y and Z are independent, their joint density function is the product of (1) and (2), i.e.,

$$\frac{1}{\sqrt{2\pi}\, 2^{v/2}\Gamma(v/2)} z^{(v/2)-1} e^{-(y^2+z)/2}$$

for $-\infty < y < +\infty$, $z > 0$.

The distribution function of $T = Y/\sqrt{Z/v}$ is

$$\begin{aligned} F(x) &= P(T \le x) = P(Y \le x\sqrt{Z/v}) \\ &= \frac{1}{\sqrt{2\pi}2^{v/2}\Gamma(v/2)} \iint_{\mathcal{R}} z^{(v/2)-1} e^{-(y^2+z)/2}\, dy\, dz \end{aligned}$$

where the integral is taken over the region $\mathcal{R}$ of the yz plane for which $y \le x\sqrt{z/v}$. We first fix z and integrate with respect to y from $-\infty$ to $x\sqrt{z/v}$. Then we integrate with respect to z from 0 to ∞. We therefore have

$$F(x) = \frac{1}{\sqrt{2\pi}2^{v/2}\Gamma(v/2)} \int_{z=0}^{\infty} z^{(v/2)-1} e^{-z/2} \left[\int_{y=-\infty}^{x\sqrt{z/v}} e^{-y^2/2}\, dy \right] dz$$

Letting $y = u\sqrt{z/v}$ in the bracketed integral, we find

$$F(x) = \frac{1}{\sqrt{2\pi}\,2^{v/2}\Gamma(v/2)}\int_{z=0}^{\infty}\int_{u=-\infty}^{\infty} z^{(v/2)-1}e^{-z/2}\sqrt{z/v}\,e^{-u^2z/2v}\,du\,dz$$

$$= \frac{1}{\sqrt{2\pi v}\,2^{v/2}\Gamma(v/2)}\int_{u=-\infty}^{x}\left[\int_{z=0}^{\infty} z^{(v-1)/2}\,e^{-(z/2)[1+(u^2/v)]}\,dz\right]du$$

Letting $w = \frac{z}{2}\left(1 + \frac{u^2}{v}\right)$, this can then be written

$$F(x) = \frac{1}{\sqrt{2\pi v}\,2^{v/2}\Gamma(v/2)}\cdot 2^{(v+1)/2}\int_{u=-\infty}^{x}\left[\int_{w=0}^{\infty}\frac{w^{(v-1)/2}e^{-w}}{(1 + u^2/v)^{(v+1)/2}}\,dw\right]du$$

$$= \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v}\,\Gamma\left(\frac{v}{2}\right)}\int_{u=-\infty}^{x}\frac{du}{(1 + u^2/v)^{(v+1)/2}}$$

as required.

4.43. The graph of Student's t distribution with 9 degrees of freedom is shown in Fig. 4-19. Find the value of t_1 for which

(a) the shaded area on the right = 0.05,

(b) the total shaded area = 0.05,

(c) the total unshaded area = 0.99,

(d) the shaded area on the left = 0.01,

(e) the area to the left of t_1 is 0.90.

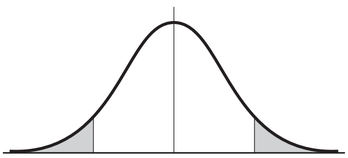


Fig. 4-19

(a) If the shaded area on the right is 0.05, then the area to the left of t_1 is $(1 - 0.05) = 0.95$, and t_1 represents the 95th percentile, $t_{0.95}$.

Referring to the table in Appendix D, proceed downward under the column headed v until entry 9 is reached. Then proceed right to the column headed $t_{0.95}$. The result 1.83 is the required value of t.

(b) If the total shaded area is 0.05, then the shaded area on the right is 0.025 by symmetry. Therefore, the area to the left of t_1 is $(1 - 0.025) = 0.975$, and t_1 represents the 97.5th percentile, $t_{0.975}$. From Appendix D, we find 2.26 as the required value of t.

(c) If the total unshaded area is 0.99, then the total shaded area is $(1 - 0.99) = 0.01$, and the shaded area to the right is $0.01/2 = 0.005$. From the table we find $t_{0.995} = 3.25$.

(d) If the shaded area on the left is 0.01, then by symmetry the shaded area on the right is 0.01. From the table, $t_{0.99} = 2.82$. Therefore, the value of t for which the shaded area on the left is 0.01 is -2.82.

(e) If the area to the left of t_1 is 0.90, then t_1 corresponds to the 90th percentile, $t_{0.90}$, which from the table equals 1.38.

4.44. Find the values of t for which the area of the right-hand tail of the t distribution is 0.05 if the number of degrees of freedom ν is equal to (a) 16, (b) 27, (c) 200.

Referring to Appendix D, we find in the column headed $t_{0.95}$ the values: (a) 1.75 corresponding to $\nu = 16$; (b) 1.70 corresponding to $\nu = 27$; (c) 1.645 corresponding to $\nu = 200$. (The latter is the value that would be obtained by using the normal curve. In Appendix D this value corresponds to the entry in the last row marked ∞.)

The *F* distribution

4.45. Prove Theorem 4-7.

The joint density function of V_1 and V_2 is given by

$$f(v_1, v_2) = \left(\frac{1}{2^{\nu_1/2}\Gamma(\nu_1/2)}\, v_1^{(\nu_1/2)-1}e^{-v_1/2}\right)\left(\frac{1}{2^{\nu_2/2}\Gamma(\nu_2/2)}\, v_2^{(\nu_2/2)-1}e^{-v_2/2}\right)$$
$$= \frac{1}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)}\, v_1^{(\nu_1/2)-1}v_2^{(\nu_2/2)-1}e^{-(v_1+v_2)/2}$$

if $v_1 > 0$, $v_2 > 0$ and 0 otherwise. Make the transformation

$$u = \frac{v_1/\nu_1}{v_2/\nu_2} = \frac{\nu_2 v_1}{\nu_1 v_2}, \qquad w = v_2 \qquad \text{or} \qquad v_1 = \frac{\nu_1 u w}{\nu_2} \qquad v_2 = w$$

Then the Jacobian is

$$\frac{\partial(v_1, v_2)}{\partial(u, w)} = \begin{vmatrix} \partial v_1/\partial u & \partial v_1/\partial w \\ \partial v_2/\partial u & \partial v_2/\partial w \end{vmatrix} = \begin{vmatrix} \nu_1 w/\nu_2 & \nu_1 u/\nu_2 \\ 0 & 1 \end{vmatrix} = \frac{\nu_1 w}{\nu_2}$$

Denoting the density as a function of u and w by $g(u, w)$, we thus have

$$g(u, w) = \frac{1}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)}\left(\frac{\nu_1 u w}{\nu_2}\right)^{(\nu_1/2)-1} w^{(\nu_2/2)-1}e^{-[1+(\nu_1 u/\nu_2)](w/2)}\,\frac{\nu_1 w}{\nu_2}$$

if $u > 0$, $w > 0$ and 0 otherwise.

The (marginal) density function of U can now be found by integrating with respect to w from 0 to ∞, i.e.,

$$h(u) = \frac{(\nu_1/\nu_2)^{\nu_1/2}u^{(\nu_1/2)-1}}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)}\int_0^\infty w^{[(\nu_1+\nu_2)/2]-1}e^{-[1+(\nu_1 u/\nu_2)](w/2)}\,dw$$

if $u > 0$ and 0 if $u \le 0$. But from 15, Appendix A,

$$\int_0^\infty w^{p-1}e^{-aw}\,dw = \frac{\Gamma(p)}{a^p}$$

Therefore, we have

$$h(u) = \frac{(\nu_1/\nu_2)^{\nu_1/2}u^{(\nu_1/2)-1}\Gamma\left(\dfrac{\nu_1+\nu_2}{2}\right)}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\left[\dfrac{1}{2}\left(1+\dfrac{\nu_1 u}{\nu_2}\right)\right]^{(\nu_1+\nu_2)/2}}$$
$$= \frac{\Gamma\left(\dfrac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\dfrac{\nu_1}{2}\right)\Gamma\left(\dfrac{\nu_2}{2}\right)}\nu_1^{\nu_1/2}\nu_2^{\nu_2/2}u^{(\nu_1/2)-1}(\nu_2+\nu_1 u)^{-(\nu_1+\nu_2)/2}$$

if $u > 0$ and 0 if $u \le 0$, which is the required result.

4.46. Prove that the F distribution is unimodal at the value $\left(\frac{v_1 - 2}{v_1}\right)\left(\frac{v_2}{v_2 + 2}\right)$ if $v_1 > 2$.

The mode locates the maximum value of the density function. Apart from a constant, the density function of the F distribution is

$$u^{(v_1/2)-1}(v_2 + v_1 u)^{-(v_1+v_2)/2}$$

If this has a relative maximum, it will occur where the derivative is zero, i.e.,

$$\left(\frac{v_1}{2} - 1\right)u^{(v_1/2)-2}(v_2 + v_1 u)^{-(v_1+v_2)/2} - u^{(v_1/2)-1}v_1\left(\frac{v_1 + v_2}{2}\right)(v_2 + v_1 u)^{-[(v_1+v_2)/2]-1} = 0$$

Dividing by $u^{(v_1/2)-2}(v_2 + v_1 u)^{-[(v_1+v_2)/2]-1}$, $u \neq 0$, we find

$$\left(\frac{v_1}{2} - 1\right)(v_2 + v_1 u) - uv_1\left(\frac{v_1 + v_2}{2}\right) = 0 \qquad \text{or} \qquad u = \left(\frac{v_1 - 2}{v_1}\right)\left(\frac{v_2}{v_2 + 2}\right)$$

Using the second-derivative test, we can show that this actually gives the maximum.

4.47. Using the table for the F distribution in Appendix F, find (a) $F_{0.95,10,15}$, (b) $F_{0.99,15,9}$, (c) $F_{0.05,8,30}$, (d) $F_{0.01,15,9}$.

(a) From Appendix F, where $v_1 = 10$, $v_2 = 15$, we find $F_{0.95,10,15} = 2.54$.

(b) From Appendix F, where $v_1 = 15$, $v_2 = 9$, we find $F_{0.99,15,9} = 4.96$.

(c) By Theorem 4-8, page 117, $F_{0.05,8,30} = \dfrac{1}{F_{0.95,30,8}} = \dfrac{1}{3.08} = 0.325$.

(d) By Theorem 4-8, page 117, $F_{0.01,15,9} = \dfrac{1}{F_{0.99,9,15}} = \dfrac{1}{3.89} = 0.257$.

Relationships among F, χ^2, and t distributions

4.48. Verify that (a) $F_{0.95} = t^2_{0.975}$, (b) $F_{0.99} = t^2_{0.995}$.

(a) Compare the entries in the first column of the $F_{0.95}$ table in Appendix F with those in the t distribution under $t_{0.975}$. We see that

$$161 = (12.71)^2, \qquad 18.5 = (4.30)^2, \qquad 10.1 = (3.18)^2, \qquad 7.71 = (2.78)^2, \qquad \text{etc.}$$

(b) Compare the entries in the first column of the $F_{0.99}$ table in Appendix F with those in the t distribution under $t_{0.995}$. We see that

$$4050 = (63.66)^2, \qquad 98.5 = (9.92)^2, \qquad 34.1 = (5.84)^2, \qquad 21.2 = (4.60)^2, \qquad \text{etc.}$$

4.49. Prove Theorem 4-9, page 117, which can be briefly stated as

$$F_{1-p} = t^2_{1-(p/2)}$$

and therefore generalize the results of Problem 4.48.

Let $v_1 = 1$, $v_2 = v$ in the density function for the F distribution [(*45*), page 116]. Then

$$\begin{aligned} f(u) &= \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{v}{2}\right)} v^{v/2}u^{-1/2}(v + u)^{-(v+1)/2} \\ &= \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{v}{2}\right)} v^{v/2}u^{-1/2}v^{-(v+1)/2}\left(1 + \frac{u}{v}\right)^{-(v+1)/2} \\ &= \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)} u^{-1/2}\left(1 + \frac{u}{v}\right)^{-(v+1)/2} \end{aligned}$$

for $u > 0$, and $f(u) = 0$ for $u \leq 0$. Now, by the definition of a percentile value, F_{1-p} is the number such that $P(U \leq F_{1-p}) = 1 - p$. Therefore,

$$\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)} \int_0^{F_{1-p}} u^{-1/2}\left(1 + \frac{u}{v}\right)^{-(v+1)/2} du = 1 - p$$

In the integral make the change of variable $t = +\sqrt{u}$:

$$2\,\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)} \int_0^{+\sqrt{F_{1-p}}} \left(1 + \frac{t^2}{v}\right)^{-(v+1)/2} dt = 1 - p$$

Comparing with (42), page 115, we see that the left-hand side of the last equation equals

$$2 \cdot P(0 < T \leq +\sqrt{F_{1-p}})$$

where T is a random variable having Student's t distribution with v degrees of freedom. Therefore,

$$\begin{aligned}\frac{1-p}{2} &= P(0 < T \leq +\sqrt{F_{1-p}}) \\ &= P(T \leq +\sqrt{F_{1-p}}) - P(T \leq 0) \\ &= P(T \leq +\sqrt{F_{1-p}}) - \tfrac{1}{2}\end{aligned}$$

where we have used the symmetry of the t distribution. Solving, we have

$$P(T \leq +\sqrt{F_{1-p}}) = 1 - \frac{p}{2}$$

But, by definition, $t_{1-(p/2)}$ is the number such that

$$P(T \leq t_{1-(p/2)}) = 1 - \frac{p}{2}$$

and this number is uniquely determined, since the density function of the t distribution is strictly positive. Therefore,

$$+\sqrt{F_{1-p}} = t_{1-(p/2)} \qquad \text{or} \qquad F_{1-p} = t^2_{1-(p/2)}$$

which was to be proved.

4.50. Verify Theorem 4-10, page 117, for (a) $p = 0.95$, (b) $p = 0.99$.

(a) Compare the entries in the last row of the $F_{0.95}$ table in Appendix F (corresponding to $v_2 = \infty$) with the entries under $\chi^2_{0.95}$ in Appendix E. Then we see that

$$3.84 = \frac{3.84}{1}, \quad 3.00 = \frac{5.99}{2}, \quad 2.60 = \frac{7.81}{3}, \quad 2.37 = \frac{9.49}{4}, \quad 2.21 = \frac{11.1}{5}, \quad \text{etc.}$$

which provides the required verification.

(b) Compare the entries in the last row of the $F_{0.99}$ table in Appendix F (corresponding to $v_2 = \infty$) with the entries under $\chi^2_{0.99}$ in Appendix E. Then we see that

$$6.63 = \frac{6.63}{1}, \quad 4.61 = \frac{9.21}{2}, \quad 3.78 = \frac{11.3}{3}, \quad 3.32 = \frac{13.3}{4}, \quad 3.02 = \frac{15.1}{5}, \quad \text{etc.}$$

which provides the required verification.

The general proof of Theorem 4-10 follows by letting $v_2 \to \infty$ in the F distribution on page 116.

The bivariate normal distribution

4.51. Suppose that X and Y are random variables whose joint density function is the bivariate normal distribution. Show that X and Y are independent if and only if their correlation coefficient is zero.

If the correlation coefficient $\rho = 0$, then the bivariate normal density function (49), page 117, becomes

$$f(x, y) = \left[\frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x-\mu_1)^2/2\sigma_1^2}\right]\left[\frac{1}{\sigma_2\sqrt{2\pi}} e^{-(y-\mu_2)^2/2\sigma_2^2}\right]$$

and since this is a product of a function of x alone and a function of y alone for all values of x and y, it follows that X and Y are independent.

Conversely, if X and Y are independent, $f(x, y)$ given by (49) must for all values of x and y be the product of a function of x alone and a function of y alone. This is possible only if $\rho = 0$.

Miscellaneous distributions

4.52. Find the probability that in successive tosses of a fair die, a 3 will come up for the first time on the fifth toss.

Method 1

The probability of not getting a 3 on the first toss is 5/6. Similarly, the probability of not getting a 3 on the second toss is 5/6, etc. Then the probability of not getting a 3 on the first 4 tosses is $(5/6)\,(5/6)\,(5/6)\,(5/6) = (5/6)^4$. Therefore, since the probability of getting a 3 on the fifth toss is 1/6, the required probability is

$$\left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right) = \frac{625}{7776}$$

Method 2 (using formula)

Using the geometric distribution, page 117, with $p = 1/6$, $q = 5/6$, $x = 5$, we see that the required probability is

$$\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^4 = \frac{625}{7776}$$

4.53. Verify the expressions given for (a) the mean, (b) the variance, of the Weibull distribution, page 118.

(a)
$$\begin{aligned}\mu = E(X) &= \int_0^\infty abx^b e^{-ax^b}\,dx \\ &= \frac{ab}{a^{1/b}}\int_0^\infty \left(\frac{u}{a}\right) e^{-u}\,\frac{1}{b}\,u^{(1/b)-1}\,du \\ &= a^{-1/b}\int_0^\infty u^{1/b}e^{-u}\,du \\ &= a^{-1/b}\Gamma\left(1 + \frac{1}{b}\right)\end{aligned}$$

where we have used the substitution $u = ax^b$ to evaluate the integral.

(b)
$$\begin{aligned}E(X^2) &= \int_0^\infty abx^{b+1}e^{-ax^b}dx \\ &= \frac{ab}{a^{1/b}}\int_0^\infty \left(\frac{u}{a}\right)^{1+(1/b)} e^{-u}\,\frac{1}{b}\,u^{(1/b)-1}\,du \\ &= a^{-2/b}\int_0^\infty u^{2/b}\,e^{-u}\,du \\ &= a^{-2/b}\Gamma\left(1 + \frac{2}{b}\right)\end{aligned}$$

Then

$$\begin{aligned}\sigma^2 &= E[(X - \mu)^2] = E(X^2) - \mu^2 \\ &= a^{-2/b}\left[\Gamma\left(1 + \frac{2}{b}\right) - \Gamma^2\left(1 + \frac{1}{b}\right)\right]\end{aligned}$$

Miscellaneous problems

4.54. The probability that an entering college student will graduate is 0.4. Determine the probability that out of 5 students (a) none, (b) 1, (c) at least 1, will graduate.

(a) $P(\text{none will graduate}) = {}_5C_0(0.4)^0(0.6)^5 = 0.07776$, or about 0.08

(b) $P(\text{1 will graduate}) = {}_5C_1(0.4)^1(0.6)^4 = 0.2592$, or about 0.26

(c) $P(\text{at least 1 will graduate}) = 1 - P(\text{none will graduate}) = 0.92224$, or about 0.92

4.55. What is the probability of getting a total of 9 (a) twice, (b) at least twice in 6 tosses of a pair of dice?

Each of the 6 ways in which the first die can fall can be associated with each of the 6 ways in which the second die can fall, so there are $6 \cdot 6 = 36$ ways in which both dice can fall. These are: 1 on the first die and 1 on the second die, 1 on the first die and 2 on the second die, etc., denoted by (1, 1), (1, 2), etc.

Of these 36 ways, all equally likely if the dice are fair, a total of 9 occurs in 4 cases: (3, 6), (4, 5), (5, 4), (6, 3). Then the probability of a total of 9 in a single toss of a pair of dice is $p = 4/36 = 1/9$, and the probability of not getting a total of 9 in a single toss of a pair of dice is $q = 1 - p = 8/9$.

(a) $$P(\text{two 9s in 6 tosses}) = {}_6C_2\left(\frac{1}{9}\right)^2\left(\frac{8}{9}\right)^{6-2} = \frac{61{,}440}{531{,}441}$$

(b) $$P(\text{at least two 9s}) = P(\text{two 9s}) + P(\text{three 9s}) + P(\text{four 9s}) + P(\text{five 9s}) + P(\text{six 9s})$$

$$= {}_6C_2\left(\frac{1}{9}\right)^2\left(\frac{8}{9}\right)^4 + {}_6C_3\left(\frac{1}{9}\right)^3\left(\frac{8}{9}\right)^3 + {}_6C_4\left(\frac{1}{9}\right)^4\left(\frac{8}{9}\right)^2 + {}_6C_5\left(\frac{1}{9}\right)^5\frac{8}{9} + {}_6C_6\left(\frac{1}{9}\right)^2$$

$$= \frac{61{,}440}{531{,}441} + \frac{10{,}240}{531{,}441} + \frac{960}{531{,}441} + \frac{48}{531{,}441} + \frac{1}{531{,}441} = \frac{72{,}689}{531{,}441}$$

Another method

$$P(\text{at least two 9s}) = 1 - P(\text{zero 9s}) - P(\text{one 9})$$

$$= 1 - {}_6C_0\left(\frac{1}{9}\right)^0\left(\frac{8}{9}\right)^6 - {}_6C_1\left(\frac{1}{9}\right)^1\left(\frac{8}{9}\right)^5 = \frac{72{,}689}{531{,}441}$$

4.56. If the probability of a defective bolt is 0.1, find (a) the mean, (b) the standard deviation for the distribution of defective bolts in a total of 400.

(a) Mean $= np = 400(0.1) = 40$, i.e., we can *expect* 40 bolts to be defective.

(b) Variance $= npq = 400(0.1)(0.9) = 36$. Hence the standard deviation $= \sqrt{36} = 6$.

4.57. Find the coefficients of (a) skewness, (b) kurtosis of the distribution in Problem 4.56.

(a) $$\text{Coefficient of skewness} = \frac{q - p}{\sqrt{npq}} = \frac{0.9 - 0.1}{6} = 0.133$$

Since this is positive, the distribution is skewed to the right.

(b) $$\text{Coefficient of kurtosis} = 3 + \frac{1 - 6pq}{npq} = 3 + \frac{1 - 6(0.1)(0.9)}{36} = 3.01$$

The distribution is slightly more peaked than the normal distribution.

4.58. The grades on a short quiz in biology were 0, 1, 2, . . . , 10 points, depending on the number answered correctly out of 10 questions. The mean grade was 6.7, and the standard deviation was 1.2. Assuming the grades to be normally distributed, determine (a) the percentage of students scoring 6 points, (b) the maximum grade of the lowest 10% of the class, (c) the minimum grade of the highest 10% of the class.

(a) To apply the normal distribution to discrete data, it is necessary to treat the data as if they were continuous. Thus a score of 6 points is considered as 5.5 to 6.5 points. See Fig. 4-20.

$$5.5 \text{ in standard units} = (5.5 - 6.7)/1.2 = -1.0$$

$$6.5 \text{ in standard units} = (6.5 - 6.7)/1.2 = -0.17$$

$$\begin{aligned}\text{Required proportion} &= \text{area between } z = -1 \text{ and } z = -0.17\\ &= (\text{area between } z = -1 \text{ and } z = 0)\\ &\quad - (\text{area between } z = -0.17 \text{ and } z = 0)\\ &= 0.3413 - 0.0675 = 0.2738 = 27\%\end{aligned}$$

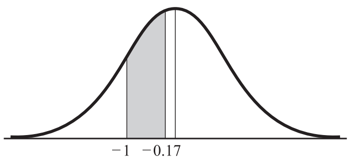


Fig. 4-20

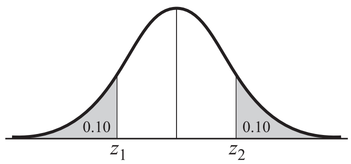


Fig. 4-21

(b) Let x_1 be the required maximum grade and z_1 its equivalent in standard units. From Fig. 4-21 the area to the left of z_1 is 10% = 0.10; hence,

$$\text{Area between } z_1 \text{ and } 0 = 0.40$$

and $z_1 = -1.28$ (very closely).
Then $z_1 = (x_1 - 6.7)/1.2 = -1.28$ and $x_1 = 5.2$ or 5 to the nearest integer.

(c) Let x_2 be the required minimum grade and z_2 the same grade in standard units. From (b), by symmetry, $z_2 = 1.28$. Then $(x_2 - 6.7)/1.2 = 1.28$, and $x_2 = 8.2$ or 8 to the nearest integer.

4.59. A Geiger counter is used to count the arrivals of radioactive particles. Find the probability that in time t no particles will be counted.

Let Fig. 4-22 represent the time axis with O as the origin. The probability that a particle is counted in a small time Δt is proportional to Δt and so can be written as $\lambda \Delta t$. Therefore, the probability of no count in time Δt is $1 - \lambda \Delta t$. More precisely, there will be additional terms involving $(\Delta t)^2$ and higher orders, but these are negligible if Δt is small.

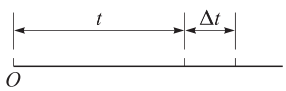


Fig. 4-22

Let $P_0(t)$ be the probability of no count in time t. Then $P_0(t + \Delta t)$ is the probability of no count in time $t + \Delta t$. If the arrivals of the particles are assumed to be independent events, the probability of no count in time $t + \Delta t$ is the product of the probability of no count in time t and the probability of no count in time Δt. Therefore, neglecting terms involving $(\Delta t)^2$ and higher, we have

(1) $$P_0(t + \Delta t) = P_0(t)[1 - \lambda \Delta t]$$

From (1) we obtain

(2) $$\lim_{\Delta t \to 0} \frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = -\lambda P_0(t)$$

i.e.,

(3) $$\frac{dP_0}{dt} = -\lambda P_0 \qquad \text{or} \qquad \frac{dP_0}{P_0} = -\lambda\, dt$$

Solving (3) by integration we obtain

$$\ln P_0 = -\lambda t + c_1 \qquad \text{or} \qquad P_0(t) = ce^{-\lambda t}$$

To determine c, note that if $t = 0$, $P_0(0) = c$ is the probability of no counts in time zero, which is of course 1. Thus $c = 1$ and the required probability is

(4) $$P_0(t) = e^{-\lambda t}$$

4.60. Referring to Problem 4.59, find the probability of exactly one count in time t.

Let $P_1(t)$ be the probability of 1 count in time t, so that $P_1(t + \Delta t)$ is the probability of 1 count in time $t + \Delta t$. Now we will have 1 count in time $t + \Delta t$ in the following two mutually exclusive cases:

(i) 1 count in time t and 0 counts in time Δt

(ii) 0 counts in time t and 1 count in time Δt

The probability of (i) is $P_1(t)(1 - \lambda\Delta t)$.
The probability of (ii) is $P_0(t)\,\lambda\Delta t$.
Thus, apart from terms involving $(\Delta t)^2$ and higher,

(1) $$P_1(t + \Delta t) = P_1(t)(1 - \lambda\Delta t) + P_0(t)\lambda\Delta t$$

This can be written

(2) $$\frac{P_1(t + \Delta t) - P_1(t)}{\Delta t} = \lambda P_0(t) - \lambda P_1(t)$$

Taking the limit as $\Delta t \to 0$ and using the expression for $P_0(t)$ obtained in Problem 4.59, this becomes

(3) $$\frac{dP_1}{dt} = \lambda e^{-\lambda t} - \lambda P_1$$

or

(4) $$\frac{dP_1}{dt} + \lambda P_1 = \lambda e^{-\lambda t}$$

Multiplying by $e^{\lambda t}$, this can be written

(5) $$\frac{d}{dt}(e^{\lambda t}P_1) = \lambda$$

which yields on integrating

(6) $$P_1(t) = \lambda t e^{-\lambda t} + c_2 e^{-\lambda t}$$

If $t = 0$, $P_1(0)$ is the probability of 1 count in time 0, which is zero. Using this in (6), we find $c_2 = 0$. Therefore,

(7) $$P_1(t) = \lambda t e^{-\lambda t}$$

By continuing in this manner, we can show that the probability of exactly n counts in time t is given by

(8) $$P_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

which is the Poisson distribution.

SUPPLEMENTARY PROBLEMS

The binomial distribution

4.61. Find the probability that in tossing a fair coin 6 times, there will appear (a) 0, (b) 1, (c) 2, (d) 3, (e) 4, (f) 5, (g) 6 heads.

4.62. Find the probability of (a) 2 or more heads, (b) fewer than 4 heads, in a single toss of 6 fair coins.

4.63. If X denotes the number of heads in a single toss of 4 fair coins, find (a) $P(X = 3)$, (b) $P(X < 2)$, (c) $P(X \leq 2)$, (d) $P(1 < X \leq 3)$.

4.64. Out of 800 families with 5 children each, how many would you expect to have (a) 3 boys, (b) 5 girls, (c) either 2 or 3 boys? Assume equal probabilities for boys and girls.

4.65. Find the probability of getting a total of 11 (a) once, (b) twice, in two tosses of a pair of fair dice.

4.66. What is the probability of getting a 9 exactly once in 3 throws with a pair of dice?

4.67. Find the probability of guessing correctly at least 6 of the 10 answers on a true-false examination.

4.68. An insurance sales representative sells policies to 5 men, all of identical age and in good health. According to the actuarial tables, the probability that a man of this particular age will be alive 30 years hence is $\frac{2}{3}$. Find the probability that in 30 years (a) all 5 men, (b) at least 3 men, (c) only 2 men, (d) at least 1 man will be alive.

4.69. Compute the (a) mean, (b) standard deviation, (c) coefficient of skewness, (d) coefficient of kurtosis for a binomial distribution in which $p = 0.7$ and $n = 60$. Interpret the results.

4.70. Show that if a binomial distribution with $n = 100$ is symmetric; its coefficient of kurtosis is 2.9.

4.71. Evaluate (a) $\sum(x - \mu)^3 f(x)$, (b) $\sum(x - \mu)^4 f(x)$ the binomial distribution.

The normal distribution

4.72. On a statistics examination the mean was 78 and the standard deviation was 10. (a) Determine the standard scores of two students whose grades were 93 and 62, respectively, (b) Determine the grades of two students whose standard scores were -0.6 and 1.2, respectively.

4.73. Find (a) the mean, (b) the standard deviation on an examination in which grades of 70 and 88 correspond to standard scores of -0.6 and 1.4, respectively.

4.74. Find the area under the normal curve between (a) $z = -1.20$ and $z = 2.40$, (b) $z = 1.23$ and $z = 1.87$, (c) $z = -2.35$ and $z = -0.50$.

4.75. Find the area under the normal curve (a) to the left of $z = -1.78$, (b) to the left of $z = 0.56$, (c) to the right of $z = -1.45$, (d) corresponding to $z \geq 2.16$, (e) corresponding to $-0.80 \leq z \leq 1.53$, (f) to the left of $z = -2.52$ and to the right of $z = 1.83$.

4.76. If Z is normally distributed with mean 0 and variance 1, find: (a) $P(Z \geq -1.64)$, (b) $P(-1.96 \leq Z \leq 1.96)$, (c) $P(|Z| \geq 1)$.

4.77. Find the values of z such that (a) the area to the right of z is 0.2266, (b) the area to the left of z is 0.0314, (c) the area between -0.23 and z is 0.5722, (d) the area between 1.15 and z is 0.0730, (e) the area between $-z$ and z is 0.9000.

4.78. Find z_1 if $P(Z \geq z_1) = 0.84$, where z is normally distributed with mean 0 and variance 1.

4.79. If X is normally distributed with mean 5 and standard deviation 2, find $P(X > 8)$.

4.80. If the heights of 300 students are normally distributed with mean 68.0 inches and standard deviation 3.0 inches, how many students have heights (a) greater than 72 inches, (b) less than or equal to 64 inches, (c) between 65 and 71 inches inclusive, (d) equal to 68 inches? Assume the measurements to be recorded to the nearest inch.

4.81. If the diameters of ball bearings are normally distributed with mean 0.6140 inches and standard deviation 0.0025 inches, determine the percentage of ball bearings with diameters (a) between 0.610 and 0.618 inches inclusive, (b) greater than 0.617 inches, (c) less than 0.608 inches, (d) equal to 0.615 inches.

4.82. The mean grade on a final examination was 72, and the standard deviation was 9. The top 10% of the students are to receive A's. What is the minimum grade a student must get in order to receive an A?

4.83. If a set of measurements are normally distributed, what percentage of these differ from the mean by (a) more than half the standard deviation, (b) less than three quarters of the standard deviation?

4.84. If μ is the mean and σ is the standard deviation of a set of measurements that are normally distributed, what percentage of the measurements are (a) within the range $\mu \pm 2\sigma$ (b) outside the range $\mu \pm 1.2\sigma$ (c) greater than $\mu - 1.5\sigma$?

4.85. In Problem 4.84 find the constant a such that the percentage of the cases (a) within the range $\mu \pm a\sigma$ is 75%, (b) less than $\mu - a\sigma$ is 22%.

Normal approximation to binomial distribution

4.86. Find the probability that 200 tosses of a coin will result in (a) between 80 and 120 heads inclusive, (b) less than 90 heads, (c) less than 85 or more than 115 heads, (d) exactly 100 heads.

4.87. Find the probability that a student can guess correctly the answers to (a) 12 or more out of 20, (b) 24 or more out of 40, questions on a true-false examination.

4.88. A machine produces bolts which are 10% defective. Find the probability that in a random sample of 400 bolts produced by this machine, (a) at most 30, (b) between 30 and 50, (c) between 35 and 45, (d) 65 or more, of the bolts will be defective.

4.89. Find the probability of getting more than 25 "sevens" in 100 tosses of a pair of fair dice.

The Poisson distribution

4.90. If 3% of the electric bulbs manufactured by a company are defective, find the probability that in a sample of 100 bulbs, (a) 0, (b) 1, (c) 2, (d) 3, (e) 4, (f) 5 bulbs will be defective.

4.91. In Problem 4.90, find the probability that (a) more than 5, (b) between 1 and 3, (c) less than or equal to 2, bulbs will be defective.

4.92. A bag contains 1 red and 7 white marbles. A marble is drawn from the bag, and its color is observed. Then the marble is put back into the bag and the contents are thoroughly mixed. Using (a) the binomial distribution, (b) the Poisson approximation to the binomial distribution, find the probability that in 8 such drawings, a red ball is selected exactly 3 times.

4.93. According to the National Office of Vital Statistics of the U.S. Department of Health and Human Services, the average number of accidental drownings per year in the United States is 3.0 per 100,000 population. Find the probability that in a city of population 200,000 there will be (a) 0, (b) 2, (c) 6, (d) 8, (e) between 4 and 8, (f) fewer than 3, accidental drownings per year.

4.94. Prove that if X_1 and X_2 are independent Poisson variables with respective parameters λ_1 and λ_2, then $X_1 + X_2$ has a Poisson distribution with parameter $\lambda_1 + \lambda_2$. (*Hint*: Use the moment generating function.) Generalize the result to n variables.

Multinomial distribution

4.95. A fair die is tossed 6 times. Find the probability that (a) 1 "one", 2 "twos" and 3 "threes" will turn up, (b) each side will turn up once.

4.96. A box contains a very large number of red, white, blue, and yellow marbles in the ratio 4:3:2:1. Find the probability that in 10 drawings (a) 4 red, 3 white, 2 blue, and 1 yellow marble will be drawn, (b) 8 red and 2 yellow marbles will be drawn.

4.97. Find the probability of not getting a 1, 2, or 3 in 4 tosses of a fair die.

The hypergeometric distribution

4.98. A box contains 5 red and 10 white marbles. If 8 marbles are to be chosen at random (without replacement), determine the probability that (a) 4 will be red, (b) all will be white, (c) at least one will be red.

4.99. If 13 cards are to be chosen at random (without replacement) from an ordinary deck of 52 cards, find the probability that (a) 6 will be picture cards, (b) none will be picture cards.

4.100. Out of 60 applicants to a university, 40 are from the East. If 20 applicants are to be selected at random, find the probability that (a) 10, (b) not more than 2, will be from the East.

The uniform distribution

4.101. Let X be uniformly distributed in $-2 \leq x \leq 2$ Find (a) $P(X < 1)$, (b) $P(|X - 1| \geq \frac{1}{2})$.

4.102. Find (a) the third, (b) the fourth moment about the mean of a uniform distribution.

4.103. Determine the coefficient of (a) skewness, (b) kurtosis of a uniform distribution.

4.104. If X and Y are independent and both uniformly distributed in the interval from 0 to 1, find $P(|X - Y| \geq \frac{1}{2})$.

The Cauchy distribution

4.105. Suppose that X is Cauchy distributed according to (29), page 114, with $a = 2$. Find (a) $P(X < 2)$, (b) $P(X^2 \geq 12)$.

4.106. Prove that if X_1 and X_2 are independent and have the same Cauchy distribution, then their arithmetic mean also has this distribution.

4.107. Let X_1 and X_2 be independent and normally distributed with mean 0 and variance 1. Prove that $Y = X_1/X_2$ is Cauchy distributed.

The gamma distribution

4.108. A random variable X is gamma distributed with $\alpha = 3$, $\beta = 2$. Find (a) $P(X \leq 1)$, (b) $P(1 \leq X \leq 2)$.

The chi-square distribution

4.109. For a chi-square distribution with 12 degrees of freedom, find the value of χ_c^2 such that (a) the area to the right of χ_c^2 is 0.05, (b) the area to the left of χ_c^2 is 0.99, (c) the area to the right of χ_c^2 is 0.025.

4.110. Find the values of χ^2 for which the area of the right-hand tail of the χ^2 distribution is 0.05, if the number of degrees of freedom v is equal to (a) 8, (b) 19, (c) 28, (d) 40.

4.111. Work Problem 4.110 if the area of the right-hand tail is 0.01.

4.112. (a) Find χ_1^2 and χ_2^2 such that the area under the χ^2 distribution corresponding to $v = 20$ between χ_1^2 and χ_2^2 is 0.95, assuming equal areas to the right of χ_2^2 and left of χ_1^2. (b) Show that if the assumption of equal areas in part (a) is not made, the values χ_1^2 and χ_2^2 are not unique.

4.113. If the variable U is chi-square distributed with $v = 7$, find χ_1^2 and χ_2^2 such that (a) $P(U > \chi_2^2) = 0.025$, (b) $P(U < \chi_1^2) = 0.50$, (c) $P(\chi_1^2 \le U \le \chi_2^2) = 0.90$.

4.114. Find (a) $\chi_{0.05}^2$ and (b) $\chi_{0.95}^2$ for $v = 150$.

4.115. Find (a) $\chi_{0.025}^2$ and (b) $\chi_{0.975}^2$ for $v = 250$.

Student's *t* distribution

4.116. For a Student's distribution with 15 degrees of freedom, find the value of t_1 such that (a) the area to the right of t_1 is 0.01, (b) the area to the left of t_1 is 0.95, (c) the area to the right of t_1 is 0.10, (d) the combined area to the right of t_1 and to the left of $-t_1$ is 0.01, (e) the area between $-t_1$ and t_1 is 0.95.

4.117. Find the values of t for which the area of the right-hand tail of the t distribution is 0.01, if the number of degrees of freedom v is equal to (a) 4, (b) 12, (c) 25, (d) 60, (e) 150.

4.118. Find the values of t_1 for Student's distribution that satisfy each of the following conditions: (a) the area between $-t_1$ and t_1 is 0.90 and $v = 25$, (b) the area to the left of $-t_1$ is 0.025 and $v = 20$, (c) the combined area to the right of t_1 and left of $-t_1$ is 0.01 and $v = 5$, (d) the area to the right of t_1 is 0.55 and $v = 16$.

4.119. If a variable U has a Student's distribution with $v = 10$, find the constant c such that (a) $P(U > c) = 0.05$, (b) $P(-c \le U \le c) = 0.98$, (c) $P(U \le c) = 0.20$, (d) $P(U \ge c) = 0.90$.

The *F* distribution

4.120. Evaluate each of the following:
(a) $F_{0.95,15,12}$; (b) $F_{0.99,120,60}$; (c) $F_{0.99,60,24}$; (d) $F_{0.01,30,12}$; (e) $F_{0.05,9,20}$; (f) $F_{0.01,8,8}$.

ANSWERS TO SUPPLEMENTARY PROBLEMS

4.61. (a) $1/64$ (b) $3/32$ (c) $15/64$ (d) $5/16$ (e) $15/64$ (f) $3/32$ (g) $1/64$

4.62. (a) $57/64$ (b) $21/32$ **4.63.** (a) $1/4$ (b) $5/16$ (c) $11/16$ (d) $5/8$

4.64. (a) 250 (b) 25 (c) 500 **4.65.** (a) $17/162$ (b) $1/324$ **4.66.** $64/243$

4.67. $193/512$ **4.68.** (a) $32/243$ (b) $192/243$ (c) $40/243$ (d) $242/243$

4.69. (a) 42 (b) 3.550 (c) -0.1127 (d) 2.927

4.71. (a) $npq(q - p)$ (b) $npq(1 - 6pq) + 3n^2p^2q^2$ **4.72.** (a) 1.5, -1.6 (b) 72, 90

4.73. (a) 75.4 (b) 9 **4.74.** (a) 0.8767 (b) 0.0786 (c) 0.2991

4.75. (a) 0.0375 (b) 0.7123 (c) 0.9265 (d) 0.0154 (e) 0.7251 (f) 0.0395

4.76. (a) 0.9495 (b) 0.9500 (c) 0.6826

4.77. (a) 0.75 (b) -1.86 (c) 2.08 (d) 1.625 or 0.849 (e) ± 1.645

4.78. -0.995 **4.79.** 0.0668

4.80. (a) 20 (b) 36 (c) 227 (d) 40

4.81. (a) 93% (b) 8.1% (c) 0.47% (d) 15% **4.82.** 84

4.83. (a) 61.7% (b) 54.7% **4.84.** (a) 95.4% (b) 23.0% (c) 93.3%

4.85. (a) 1.15 (b) 0.77 **4.86.** (a) 0.9962 (b) 0.0687 (c) 0.0286 (d) 0.0558

4.87. (a) 0.2511 (b) 0.1342 **4.88.** (a) 0.0567 (b) 0.9198 (c) 0.6404 (d) 0.0079

4.89. 0.0089 **4.90.** (a) 0.04979 (b) 0.1494 (c) 0.2241 (d) 0.2241 (e) 0.1680 (f) 0.1008

4.91. (a) 0.0838 (b) 0.5976 (c) 0.4232 **4.92.** (a) 0.05610 (b) 0.06131

4.93. (a) 0.00248 (b) 0.04462 (c) 0.1607 (d) 0.1033 (e) 0.6964 (f) 0.0620

4.95. (a) 5/3888 (b) 5/324 **4.96.** (a) 0.0348 (b) 0.000295

4.97. 1/16 **4.98.** (a) 70/429 (b) 1/143 (c) 142/143

4.99. (a) $\binom{13}{6}\binom{39}{7}\Big/\binom{52}{13}$ (b) $\binom{13}{0}\binom{39}{13}\Big/\binom{52}{13}$

4.100. (a) $\binom{40}{10}\binom{20}{10}\Big/\binom{60}{20}$ (b) $[({}_{40}C_0)({}_{20}C_{20}) + ({}_{40}C_1)({}_{20}C_{19}) + ({}_{40}C_2)({}_{20}C_{18})]/{}_{60}C_{20}$

4.101. (a) 3/4 (b) 3/4 **4.102.** (a) 0 (b) $(b - a)^4/80$

4.103. (a) 0 (b) 9/5 **4.104.** 1/4

4.105. (a) 3/4 (b) 1/3 **4.108.** (a) $1 - \dfrac{13}{8\sqrt{e}}$ (b) $\dfrac{13}{8}e^{-1/2} - \dfrac{5}{2}e^{-1}$

4.109. (a) 21.0 (b) 26.2 (c) 23.3 **4.110.** (a) 15.5 (b) 30.1 (c) 41.3 (d) 55.8

4.111. (a) 20.1 (b) 36.2 (c) 48.3 (d) 63.7 **4.112.** (a) 9.59 and 34.2

4.113. (a) 16.0 (b) 6.35 (c) assuming equal areas in the two tails, $\chi_1^2 = 2.17$ and $\chi_2^2 = 14.1$

4.114. (a) 122.5 (b) 179.2 **4.115.** (a) 207.7 (b) 295.2

4.116. (a) 2.60 (b) 1.75 (c) 1.34 (d) 2.95 (e) 2.13

4.117. (a) 3.75 (b) 2.68 (c) 2.48 (d) 2.39 (e) 2.33

4.118. (a) 1.71 (b) 2.09 (c) 4.03 (d) -0.128

4.119. (a) 1.81 (b) 2.76 (c) -0.879 (d) -1.37

4.120. (a) 2.62 (b) 1.73 (c) 2.40 (d) 0.352 (e) 0.340 (f) 0.166